

RA Fraser C.M.; Barrell B.; human malaria parasite Plasmodium  
 RT "Genome Sequence of the human malaria parasite Plasmodium  
 RT falciparum";  
 RL Nature 418:498-511(2002).  
 DR EMBL; AR014841; AAN5938.1; -.  
 DR HSSP; Q93TC2; INNPI.  
 KW Hypothetical protein.  
 SQ SEQUENCE 2227 AA; 262841 MW; 6D5D4BRRR643339C CRC64;

Query Match 6.3%; Score 252.5; DB 2; Length 2227;  
 Best Local Similarity 19.1%; Pred. No. 0 0.024; Mismatches 319; Indels 329; Gaps 49;  
 Matches 190; Conservative 155; Mismatches 319; Indels 329; Gaps 49;  
 QY 10 FRTGLGCKE-----GSILKKDTGVEH-----HHDETEESKEK 42  
 Db 690 YNNNNNNKETCTSINIKHSENKYPFNKHSHVNSYMTNTLPHRNALTTSNNNNNEYEKEK 749  
 QY 43 SSIPTIBNIST----IRDPEK---DLKLU-----KKEERADDT----SB 79  
 Db 750 ER---DRNITGNNNNLVVEYNNSCIPPLKQMPIDGVRSNSINKLNVNVTORTSSVSY 806  
 QY 80 TCKRMBSEYDY-----KDDKGMIAYDGGTDLYETEKUDRIKSYIGVLSPSK 128  
 Db 807 TTKNLDENSPPMILINGIRBSCKIISNNI--NGNSIGFNSKLQDTHHSMV----- 858  
 QY 129 DCHPBPILGKISNSKAKVYGRNYKSEIKATKYDFHSKTMF-----DIYANTN--- 179  
 Db 859 -----NESIPLKMMGNNY-----IEHYDKNIFLVEKTYDTSNIHGI 900  
 QY 180 -----DIVGLAGAFDR-----LFVKDN---DOKK----- 202  
 Db 901 HENSLMLKVNLYKACTFHGSYRHNQKOMTYBEMUNINIQKKNYHYHNGTFLKPLVNTN 960  
 QY 203 -----AEIKRMPKKEKTS---EVYVSSGYVIEGLGDSLKQPKDULTRMEGKI 253  
 Db 961 VAVNBEPADINNSAQKRHLHSLSKMGYBOKSMRMEYRNKLYNNINNNNNNNNN 1013  
 QY 254 YSDSEKQOY---LLKONILRGYAL-----KVTTAEGKTDLEGNGVSKEDIAK--- 302  
 Db 1014 YDNDEYQOYNNSYCFDSDLKPMLPHQNSKLLTHSNNKOSPNGINVESEGHIANPEI 1073  
 QY 303 -----TOKANP-----N 309  
 Db 1074 KTFARNSYPIIQLGLNCPOLGTDPSNRQKHNWVVIKREYELNNKIGSTINVLRKG 1133  
 QY 310 IRLS-----ETIYADSAPNVRDERSRSTPSVLMALDGENILRYQVFEND 356  
 Db 1134 LAKISTHNGKPFSSFSNMNDNTNNTYMEGANIQDVAANNNNKESC-DNTKDMRTSINFVSRE 1192  
 QY 357 K-GB--AIDKQGNLVTDSKULVFGK-DDKEVGTGEDXENRAIKEK-GSMCLRID---TKP 408  
 Db 1193 SYGEHESLDVYQECYVKKNUKLI--NKYNDKCY---EDMNNSVNTEDDNASWMPYETBN SNP 1248  
 QY 409 VNLSMD---KVNPNPSKSNSKNTVRENPEFYLRGKLSDKGFWNE----LRYNESV--- 455  
 Db 1249 YIVDQERNMKVN---NVLTNNNSWYUDSKYDSEKEDNAENKSDILANNENIHKD 1304  
 QY 456 -----VDNYLIYGDLEHDNTDPNTKLVDGDDIMW--GMDYKANGPPPKV-TMDG 506  
 Db 1305 QKKEKIONNERBETISQADIENRNSOBEEVKEHBL-WVIMASNEERKSYISLYSDMS 1363  
 QY 507 NYLQTCYSDIPLAKAXGWPHQDLYDNUKPEVNIDPKENTSLEYADGKSVWVINDKNG 566  
 Db 1364 KRTVKTKYSDMWNVEVILANBDNLTTKVKVQLEKENMIDMVTETBENINTKTEND 1423  
 QY 567 PDEGEIQORHQHIVINGKEVTPSPDPIQIKDITKLNKIVWDFAART--TVK--EPLNKOTG 622  
 Db 1424 INEEVR-----NBQKRSINHIND----TNINHIDBYENDTYNFIDTECVHNENN 1472  
 QY 623 EYSELKPHRVVTIONKEMMSTIVSBEFDLIPVYKGEBEKGYQFDGWEISCPGKHDAG 682  
 Db 1473 MNSIQYQTPYDTRNHLVK---NNQPI-----PEEGLNLBNFBRK--v 1516

QY 683 YYINLSDD-----TPIKPVFCICBKKCBENKPTP-----DVSKCDN 720  
 Db 1517 YIENNKKDHDHKGSKTSNLTSIANTICKSERNDHNEKENTVVRKGSEKGIRRKVSKERN 1576  
 QY 721 PQVN-HSQLN-----ESHRKDQIREHRSOKS 745  
 Db 1577 BKLNKEENYINNYDROMDHRODNTKCKENDBN 1609

Search completed: February 16, 2005, 19:19:52  
 Job time : 188 secs

*This Page Blank (uspto)*

; LENGTH: 1787  
; TYPE: PRT  
; ORGANISM: KI Parasite Clone  
; US-10-415-253-2

Query Match 5.0%; Score 200.5; DB 15; Length 1787;  
Best Local Similarity 20.5%; Pred. No. 0.015; Mismatches 329; Indexes 301; Gaps 52;  
Matches 204; Conservative 160; MisMatches 329; Indexes 301; Gaps 52;  
QY 6 AES--KFRN-LGNGKKGSLKDT-----TGVEHHHNE-----ESIK 40  
Db 868 AESTTFSNILETORONTINDTEERKELEHENVSALENTOBEEKEVIVDIEEK 927  
Qy 41 EKSSTPI-----DRNSTIRD-FEN-----KDLCKLKKRFFEDFTS 78  
Db 928 EEVATLIELTVEQAREKSANTITEFLENENAVENAVENAVENLKENETVFTNLVKVB 987  
Qy 79 ET---GKMB--EVDYK----DDKG--NIAVDOGTDEYET----EKLD-- 114  
Db 988 ETVELSGESIENENENDKAFFSEBFUNVKGIOBNLT--GMFRS1ETSIVIQSERBKVLN 1044  
Qy 115 -EIKSKIGVGLSPSKDGHFRLGKISNVSKNAKV-----YGINYKSIKIATKIDFH 166  
Db 1045 ENVVISILOMIENENKEG----LNUKLENISTEGOETVTHVQMVYDVPANK--- 1097  
Qy 167 SKTMTPDLYANINDVGLAFAGDMRFLFKNDOKAECIKIRMPKIKETKS EYPVSSY 226  
Db 1098 -----DOPFGILNEBAGGL----KEMFFNLBDVKSESDVITVAKIKDOPVOKER 1146  
Qy 227 GNUTELGEDD----LSKNIKDMLTN----ESGKLYSDSRKQXLLKONILRKGYALK 277  
Db 1147 VSIIEEMBENIVDLEEBEKBEDTDKMDAVETESEISSEKETESIKOK--EKDVSLV 1203  
Qy 278 VTTNPQG-----KTDMLGNGVSK--BDYAKIQRKPNL---- 310  
Db 1204 VEYKODNIMADESVKVLKMRHELMKDAVEINDITSKJLBETOBLNRETAIDLKOMEK 1263  
Qy 311 -----RALSETTIYADSRNVEDGR--STSIVMSALDGFMNTIRYQVFTKNDGEAIDK 363  
Db 1264 LKELEKALSE----DSKETIADKOTLERYTEERHD----ITTLDBVELKOV 1309  
Qy 364 DGNLVLTDSKVLRGDKDEYTEDKFNVRKAKEDGSMRLPDTKPVNL----SM 413  
Db 1310 RBDKTEKVSDL---RDLSEDLK-----VRIKKELESELEDEYKELKTIETDILEEKKEI 1362  
Qy 414 DKVYFNPSKSNKITVYRNPEYLRLGKISDKOGPPWMLRVNYSVD----NVIYGL 465  
Db 1363 BKDHKEKFEHABERKEDRLEADTKEVS----SLXVEBRCKLERTHELKEBEVHILGDA 1417  
Qy 466 HIDNTRDFNIKLNKWD--GDIM-----DWGMKDYKANGFPDKVTDMGNYLQLGSDLN 518  
Db 1418 HKGLBLDDLE-EVDLKSILDMKGDMSLGLDMOKESLDVTTKIGERT----BSLKDVL 1473  
Qy 519 AKAVEVHYOFLYDN-----VKPEVNIDPKGNTSIEYADGSKVVFNINCR-RN 564  
Db 1474 SSALGMDEEQMKTRKAQRPKLEEVILKEEVKRPKKIT----KKKORFDIKOKEPKD 1528  
Qy 565 NGFDGRIOSGHIIYINGK----BTSFENDIKOILDKTN----IKUWKDFAR--N 609  
Db 1529 EIVVEREMKODIDEEDVEEDIBEDKVEDIDEDIDEDIGDODEVIDIIVQKERRIEK 1598  
Qy 610 TTVKEPLNKDGE VSELKPH-----RVTUTONGKEMSTIVSBED 651  
Db 1589 VRAKCKKBLKBLKVEVGSKKGVDESYMKVYQKIDKEVDEKVKALESKNDVTLVKND 1648  
Qy 652 FILPVTKGKBLKKGKQDGMWEISGFEGSKKDAVGWML-----SKDTPIKPFK 698  
Db 1649 FFSKVO-KNPFYKSKYKPAFPISAVA--PASYVVEFTPSLSSCTVIASTYLLSKVK 1705  
Qy 699 KIEEKKEEBENKP----TFDV-----SKKCONPQVNHQSOLNHSRKEDLQR 739  
Db 1706 TINKOKE--RPFYSPVFDFKKNLKHLOOMKKEKESKEKONNVIET--NKAEGKQWV 1760

OY 740 BEHSOKSDSTDVATVLDKNTISSKSTTNPNK 773  
Db 1761 TNKTEKT-----TKVDRGKVKPKRRTQSK 1786

Search completed: February 16, 2005, 19:32:38  
Job time : 147 secs

This Page Blank (uspto)

RESULT 15  
 US-09-949-016-11433  
 Sequence 11433, Application US/09949016  
 Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SEQ ID NO 11433  
 ; LENGTH: 2733  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 ; US-09-949-016-11433

Query Match 4.6%; Score 185; DB 4; Length 2733;  
 Best Local Similarity 20.0%; Pred. No. 0.0032; Mismatches 338; Indels 194; Gaps 39;  
 Matches 170; Conservative 150; MisMatches 338; Indels 194; Gaps 39;

Qy 21 LKQDGTGVEHHHORNBR-SIKEKSPTIDNTRDFTENKDARKLKEK--FREVD 75  
 Db 1319 LKERIAGLUEBEBKOKNKEPQSOTLENENKNTLQISI-KOBRKIQMQLQEVTYKMLAQIOE 1377  
 Qy 76 FTSEUTGKMEVYDKYDD-----KGNTIAY-DIGTDLAYETEKLDBIKSYG 122  
 Db 1378 ELSRVTKLKEBAEERKDLAERLAKNOLASLNGSIGNYCQDTDAQIKNBLIESMRNLCK 1437  
 Qy 123 VLSSKDGHRFLGKLNUSKNAKTY-----GM-----NYKSIKIA 160  
 Db 1438 CVSLLEBEBKQOLVKRKTYSBIRKYLEKIQAGQKPERKSHARKELOBLIKERQEVHQ 1497  
 Qy 161 TYKD---FHSKTMFDLYANINDVGLAFAGDMRLFVNDOKKABR-KLMPERKIST 216  
 Db 1498 LQKOCIRYQBKISALERTYALE-----FVQTESQDLEITKRNLAQAEHR 1544  
 Qy 217 KSEPYVSSYGVNTRIGBEGDLSKNUPDNLTKMESGKIVYSDSEKOOYLKDNITILRK--- 272  
 Db 1545 KRAQABELASFRVLADTDOSAARVADNLKAKELQSNKESVSKOMKQBDLTERLQA 1604  
 Qy 273 -GYALKVTTYNGPKTUDMELANGVYSKDIAKTIQKANPNLALASSETIYADSRNEDGRST 331  
 Db 1605 BEKULKEKKNMOKBLDALRKRKVKHLBTG1Q-----VTLNKDDKEYQO--- 1649  
 Qy 332 OSVIMSLADGRNLRQVTFKMDKGALDDKGNLVTSKAVLGKODKBYTGDKEN 391  
 Db 1650 ---LQENLIS-TVUQLAFTKMS---SLODDPRDVRVBAK----KOTERKFSDAIQSK 1636  
 Qy 392 VEAT---KEDG-SMPLDITKVNLSMDKNYTPNSK---SMKTY--VRNPSPYLRGKISDK- 442  
 Db 1697 BEERLKEMCNSVJQDQRMISIMBEKLNIISLHDKQIWMESKACTVWQLOQCVDT 1756  
 Qy 443 GFYWMLRIVNESSVUNLVLYTDGLHDNTDRFFNKLKVQDGIMWGMKOKYKANGEPDKYT 502  
 Db 1757 QGENKELL-----SOLARTR----HLYHSSNELAKLESLKSLSKDLT 1796  
 Qy 503 DMD-----GRV--VLTGTSSDL-YAKAVGHYQELYDNTPEV----- 537  
 Db 1797 DLNSLNLCKCKEOKGNLLEGITRQEOADIONSK-----FSYBOLRTDLSRSLTRH 1849  
 Qy 538 NIDKGNTSIEBYACK--SVVFNINDKRNFGDSRIQOBHIIYINGKEYTSP--NDIKQT 592  
 Db 1850 BINMKQKJLJSLSGKERAQVAAELRQ-HDKEKJBLNLLSQBERENIVLBRNKA 1908

Oy 593 IDKTNANIKITVKEDFARNTTWKFPLANKDTCGEVSILK--PHRVVTIIONGKEMSTIVSEE 650  
 Db 1909 VDKTNQLMETAKTICKTKEENTOOKAQDLSFVTSMSLQNDRURIVDYOQLERHLSILEK 1968  
 Qy 651 DPITLEVYKGSLBKGSQPDGEWEISGPEG-----KODAGYVIMLSKDT 691  
 Db 1969 DOLTOBAAEENKKE---BIRGLASHMDLNSENAKDABLIQYREDIANQVTT-KDS 2023  
 Qy 692 FIKPWFK-KIRBKEGERENKEPFDVSKKONPOVHSAMES----HRIKEDLQREHRSOK 745  
 Db 2024 QOKOLIKVQDQNKELNKVAKBLERKESBEEAN-BDRRSFSNALQEBEKOOLSKERLSK 2082  
 Qy 746 ---SDSTDQVTA 754  
 Db 2083 VSISQDTRQVTA 2094

Search completed: February 16, 2005, 19:20:41  
 Job time : 50 SECs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model  
Run on: February 16, 2005, 19:06:53 ; Search time 170 Seconds

(Without alignments)  
1758.622 Million cell updates/sec

Title: US-10-067-385-8  
Perfect score: 4026  
Sequence: 1 KLGIAESKEKNLNGKEGS . . . . . ATVLDKONITSSKSTTNNPNK 73

**Scoring table:**

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

**Post-processing:** Minimum Match 0% Maximum Match 10% Listing first 45

Database : A\_Geneseq\_16Dec04:\*

```
2: geneseqp1990B:*
3: geneseqp2000B:*
```

```
5 : geneseqp2002b :: *
6 : geneseqp2003ab :: *
```

7:  
8: geneseqp2003bb : \*  
8: geneseqp2004B : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
------------	-------	-------------	--------	----	----	-------------

1 4026 100.0 773 4 AAB48343 Aab48343 S. pneu  
 2 4026 100.0 2140 6 ABU01020 ABU01020 S. pneu

Abn45746 protein S  
Abn92113 S protein  
Abn92113 S protein  
Abn92113 S protein  
Abn92113 S protein

Adk48759 Stcrgtcc  
Adk48759 Novel 99  
Adk48759 Novel 99  
Adk48759 Novel 99  
Adk48759 Novel 99

9 6.5 15.3 117 2 AAW55096  
10 6.15 15.3 117 5 ABP54590  
11 6.17 15.3 117 5 ABP54590 S. pneu

12	227	5.5	1196	6	AB024813
13	222	5.5	861	7	AB023608

14	220	5.3	1100	8	AKX9180
15	218	5.4	1639	2	AKX9145
15	218	5.4	1639	2	AKX9145_P
16	218	5.4	1639	5	AaE29345

17 217 5.4 1166 8 ADK9185  
18 217 5.4 1166 8 ADK9178  
19 217 5.4 1166 8 ADK9178  
20 217 5.4 1166 8 ADK9178

20 216 5.4 1166 8 ADK99183 Ad99183  
21 216 5.4 1233 5 ABP30203 Ad99203  
22 216 5.4 1233 5 ABP29675 Str99675

23	216	5.4	1233	8	ADK9116	ADK9116	StrepCcc
24	216	5.4	1233	8	AOD10499	AOD1469	Group E
25	216	5.4	1239	8	ABP25822	ABP28822	StrepCC

it is media; nasopharyngeal; response; immunotherapy;	Aar07475	Merozoite
	Aab18179	Streptococcus
	Aab18172	Plasmodium
	Abu42127	Protein e
	Abu18183	Plasmodium
	Acp86452	Clostridium
	Aqp73658	C. perfringens
	Aab18324	Plasmodium
	Aqp25433	Plasmodium
	Abu48941	Protein e
	Abu24124	Protein e
	Akp50777	Sequence
	Aab18171	Plasmodium
	Acp25444	Plasmodium
	Aab18195	Plasmodium
	Aab18272	Plasmodium
	Abj18982	Pathogen
	Aab18244	Plasmodium
	Aay08643	S. aureus

Aad9917	Streptococcus
Aab51872	Plasmodium e
Aab54327	Protein e
Aab1818	Plasmodium
Adb64542	Clostridium
Aad973658	<i>C. perfringens</i>
Aab18324	Plasmodium
Adb65433	Plasmodium
Abo48949	Protein e
Abu24124	Protein e
Aab50777	Sequence
Aab18171	Plasmodium
Adp25444	Plasmodium
Aab18195	Plasmodium
Aab18272	Plasmodium
Abl18982	Pathogen
Aab18244	Plasmodium
Aav08643	<i>S. aureus</i>

CC present sequence represents the *S. pneumoniae* Sp130 polypeptide  
 XX  
 SQ Sequence 773 AA;

Query Match 100.0%; Score 4026; DB 4; Length 773;  
 Best Local Similarity 100.0%; Pred. No. 7. 5e-225; Mismatches 0; Indels 0; Gaps 0;  
 Matches 773; Conservative 0; Misnmatches 0; Indels 0; Gaps 0;

QY 1 KLGELIAESKPKNIGNGKEGSSLKDGTGVEHHHOENNESEIKENSSFTIDRNISTIRDFENK 60  
 Db 1 KLGEBIASSESKPKNIGNGKEGSSLKDGTGVEHHHOENNESEIKENSSFTIDRNISTIRDFENK 60  
 QY 61 DLKKLIKKEKPFREVDIFSETGRMEVYDYYKDDKGNIAYDDGTDLSEYETEKLDEJKSKI 120  
 Db 61 DLKKLIKKEKPFREVDIFSETGRMEVYDYYKDDKGNIAYDDGTDLSEYETEKLDEJKSKI 120  
 QY 121 YGTLSPSKDGFHETILGKISNVNSNAKVVYGGNNKSTIEKATKDYFPIKSTMTMDLYANIND 180  
 Db 121 YGVLSPSKDGFHETILGKISNVNSNAKVVYGGNNKSTIEKATKDYFPIKSTMTMDLYANIND 180  
 QY 181 IVGLAFAGDMRLPUKNDQKKAERKURMRPEKIKETSEYPTVSSKGWVLEGEGLDSK 240  
 Db 181 IVGLAFAGDMRLPUKNDQKKAERKURMRPEKIKETSEYPTVSSKGWVLEGEGLDSK 240  
 QY 241 KPNLNTKESGKYLSDSRQQYLAKONTLRKQYALKPTVNGKTMLENGVSYED 300  
 Db 241 KPDNLTKESGKYLSDSRQQYLAKONTLRKQYALKPTVNGKTMLENGVSYED 300  
 QY 301 AKTQKANPNTLRAUSETTYADSRSNRVEDGRSTOSVYALDGENIIRYQVFTEKNDKGEA 360  
 Db 301 AKTQKANPNTLRAUSETTYADSRSNRVEDGRSTOSVYALDGENIIRYQVFTEKNDKGEA 360  
 QY 361 IDKDGNYUTPDSKVLFLCKDKDKEYTGEDKPKVTEAIKEDGSMFLIDTKEVNLMSMDKMYFNP 420  
 Db 361 IDKDGNYUTPDSKVLFLCKDKDKEYTGEDKPKVTEAIKEDGSMFLIDTKEVNLMSMDKMYFNP 420  
 QY 421 SKANKITTPRPLRKGISDKGPNHLRVMESVWNLTYIGDHDNTDNPNIKAVK 480  
 Db 421 SKANKITTPRPLRKGISDKGPNHLRVMESVWNLTYIGDHDNTDNPNIKAVK 480  
 QY 481 DGDDIMDMNKMDYKANGFPDKVUDMGNYLTQGYSIDLNAKAVGVHVKPYDNVKPEVNTD 540  
 Db 481 DGDDIMDMNKMDYKANGFPDKVUDMGNYLTQGYSIDLNAKAVGVHVKPYDNVKPEVNTD 540  
 QY 541 PKGNTSTYEADGSKSVVENINDKRNFGDFOBQHTYINGKEYTSENDIKQDNTK 600  
 Db 541 PKGNTSTYEADGSKSVVENINDKRNFGDFOBQHTYINGKEYTSENDIKQDNTK 600  
 QY 601 IIVKDRANTTKEPLANKTGESELKPRHTITVONGKEMSTTISEBBLIPVKE 660  
 Db 601 IIVKDRANTTKEPLANKTGESELKPRHTITVONGKEMSTTISEBBLIPVKE 660  
 QY 661 LKGYQFDGWETSGFEGKDDAGYVINTSKDTPTKPVFKRKEERKEENKPTFDVSKCRDN 720  
 Db 661 LKGYQFDGWETSGFEGKDDAGYVINTSKDTPTKPVFKRKEERKEENKPTFDVSKCRDN 720  
 QY 721 PONHSOLNESEIRKEDQREIISQKSSTKDYTAVIDKNISSKSTNNPK 773  
 Db 721 PONHSOLNESEIRKEDQREIISQKSSTKDYTAVIDKNISSKSTNNPK 773  
 XX

## RESULT 2

ABU01020 ID ABU01020 standard; protein; 2140 AA.  
 XX  
 AC ABU01020;  
 XX  
 DT 23-OCT-2003 (revised)  
 11-FEB-2003 (first entry)

S. pneumoniae type 4 strain protein from coding region #590.

Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;

KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
 KW gene therapy; vaccine.  
 XX Streptococcus pneumoniae; type 4 strain.

XX  
 OS PD 03-OCT-2002.  
 XX PP 27-MAR-2002; 2002WO-1B002163.  
 XX PR 27-MAR-2001; 2001GB-00007658.  
 XX PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 PI Masignani V, Tettelin H, Fraser C;  
 XX DR WPI; 2003-040579/03.  
 XX N-PSDB; ABX06302.

PS Claim 1; SEQ ID NO 1180; 56pp; English.  
 XX  
 CC New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,  
 CC useful as medicaments for treating or preventing a disease or infection  
 CC due to *streptococcus* bacteria, such as pneumonia, sepsis, otitis media or  
 CC ear infection.

CC The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
*Streptococcus pneumoniae* type 4 strain genomic sequence appearing as  
 ABS5654. Also included are an antibody which binds one of the proteins,  
 CC treating a patient by administering the protein, DNA or antibody (in a  
 CC composition), a kit comprising first and second primers, which are the  
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
 CC sequence not defined in the specification, for amplifying a target  
 CC sequence contained within a *Streptococcus* nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a *Streptococcus pneumoniae* bacterium, where one or more genes  
 CC encoding the protein has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to *Streptococcus*  
 CC bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2469 proteins  
 CC expressed by the identified coding regions from the genomic sequence.  
 Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC Standardise OS field)  
 SQ sequence 2140 AA;

Query Match 100.0%; Score 4026; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 2. 9e-24; Mismatches 0; Indels 0; Gaps 0;  
 Matches 773; Conservative 0; Misnmatches 0; Indels 0; Gaps 0;

QY 1 KLGEBIASSESKPKNIGNGKEGSSLKDGTGVEHHHOENNESEIKENSSFTIDRNISTIRDFENK 60  
 Db 1334 KLGEBIASSESKPKNIGNGKEGSSLKDGTGVEHHHOENNESEIKENSSFTIDRNISTIRDFENK 1393  
 QY 61 DLKKLIKKEKPFREVDIFSETGRMEVYDYYKDDKGNIAYDDGTDLSEYETEKLDEJKSKI 120  
 Db 1394 DLKKLIKKEKPFREVDIFSETGRMEVYDYYKDDKGNIAYDDGTDLSEYETEKLDEJKSKI 1453

QY	121	YGVLSPSKDGHFEILGKISNVSKNAKYIGNNYSIEIKATKDIFHSKTMFDLYANIND	PI	Wang L,
AC	1454	YGVLSPSKDGHFEILGKISNVSKNAKYIGNNYSIEIKATKDIFHSKTMFDLYANIND	PI	Zamudio C,
XX	1454	YGVLSPSKDGHFEILGKISNVSKNAKYIGNNYSIEIKATKDIFHSKTMFDLYANIND	XX	Malone C,
DR			DR	Hanselbeck R,
QY	181	IVDGLAFAGDMRLFVKONDQKAEIKRMPKEKIKTSEPKVYVSSYGNVIELGEGLISK	DR	Ohlsen KU,
DB	1514	IVDGLAFAGDMRLFVKONDQKAEIKRMPKEKIKTSEPKVYVSSYGNVIELGEGLISK	N-PSDB;	Zyskind JW;
PR	241	KPDNLTKMESGKLYSDKQYLKONILRKGYALKVTTNPKGKTMFDLYANIND	PT	Watt D,
XX	241	KPDNLTKMESGKLYSDKQYLKONILRKGYALKVTTNPKGKTMFDLYANIND	XX	Trawick JD,
PR	1574	KPDNLTKMESGKLYSDKQYLKONILRKGYALKVTTNPKGKTMFDLYANIND	DR	Carr GJ,
PA	1574	KPDNLTKMESGKLYSDKQYLKONILRKGYALKVTTNPKGKTMFDLYANIND	N-PSDB;	Yamamoto R,
QY	301	AKIQKANPLRASETTYIADSRNVEDGRSTOSVIMSLALGENITRYQVFPMKNDGEA	PT	Forbyth RA,
DB	1634	AKIQKANPLRASETTYIADSRNVEDGRSTOSVIMSLALGENITRYQVFPMKNDGEA	XX	Xu HH;
QY	361	IDKDGNLVTDSSKLVLFLRGKDKEYTGEDKPNVAAEKGSMFLIDTPVNLSMDKNYFP	PR	WPI;
DB	1694	IDKDGNLVTDSSKLVLFLRGKDKEYTGEDKPNVAAEKGSMFLIDTPVNLSMDKNYFP	XX	2003-02926/02.
QY	421	SKSNKIYVRNPEFYLRGKISDKCGFWNLRLVNESSVNDNTLYVGDLHIDNTDPNFKLNK	XX	ACA49616.
DB	1754	SKSNKIYVRNPEFYLRGKISDKCGFWNLRLVNESSVNDNTLYVGDLHIDNTDPNFKLNK	XX	New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.
QY	481	DGDIMDKMKDYANGPDKVTDMDGNVTLQGYSIDNAKAVGWHYQFLDYNKPENID	CC	Claim 25; SEQ ID NO 73670; 1766pp; English.
DB	1814	DGDIMDKMKDYANGPDKVTDMDGNVTLQGYSIDNAKAVGWHYQFLDYNKPENID	CC	The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are:
QY	541	PKGNTSRYADGKSVFVFINDKRNNGFGBIOHQHIXINGKEYTSFNDIKQDKNIK	CC	(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than <i>S. aureus</i> , <i>S. typhimurium</i> , <i>K. pneumoniae</i> or <i>P. aeruginosa</i> . The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
DB	1974	PKGNTSRYADGKSVFVFINDKRNNGFGBIOHQHIXINGKEYTSFNDIKQDKNIK	CC	CC
QY	601	IIVKDFARNTTVKEFLINKDTGEVSELKPHRVVTIONGKMSSTIVSEBDLPLVNGE	CC	CC
DB	1934	IIVKDFARNTTVKEFLINKDTGEVSELKPHRVVTIONGKMSSTIVSEBDLPLVNGE	CC	CC
QY	661	LEKGYQFDGWEISGPEGKDDAGYVNLUSKTFIKPVFKKEEEKKPTEVDLPLVNGE	CC	CC
DB	1994	LEKGYQFDGWEISGPEGKDDAGYVNLUSKTFIKPVFKKEEEKKPTEVDLPLVNGE	CC	CC
QY	721	PQNHSQLNEHARKEDLOREBHQKSSTDKVATVLDKNNSKSTNNPK	CC	CC
DB	2054	PQNHSQLNEHARKEDLOREBHQKSSTDKVATVLDKNNSKSTNNPK	CC	CC
RESULT 3				
ID	ABU45746	standard; protein: 2140 AA.	SQ	Sequence 2140 AA;
XX	ABU45746;			XX
AC			Query Match	100.0%; Score 4026; DB 6; Length 2140;
XX			Best Local Similarity	100.0%; Pred. No. 2.9e-224;
DT	19-JUN-2003 (first entry)		Matches	773; Consistency 0; Mismatches 0; Index 0; Gaps 0;
DE	Protein encoded by Prokaryotic essential gene #31273.		QY	1 KUGEIABKFKLNGNGKESLKKDTQVEHHQNEESEIKERKSFTIDRNISTIRDEFENK 60
XX	Antisense; prokaryotic essential gene; cell proliferation; drug design.		DB	1334 KUGEIABKFKLNGNGKESLKKDTQVEHHQNEESEIKERKSFTIDRNISTIRDEFENK 1393
OS	<i>Streptococcus pneumoniae</i> .		QY	61 DLCKLKIKKFREVDFFSETGKMEBEYDVKDKGNITAYDGTDLIYEKEKIDEISKI 120
XX	W0200277183-A2.		DB	1394 DLCKLKIKKFREVDFFSETGKMEBEYDVKDKGNITAYDGTDLIYEKEKIDEISKI 1453
PR	03-OCT-2002.		QY	121 YGVLSPSKDGHFEILGKISNVSKNAKYIGNNYSIEIKATKDIFHSKTMFDLYANIND 180
XX	21-MAR-2002; 2002WO-US009107.		DB	1514 YGVLSPSKDGHFEILGKISNVSKNAKYIGNNYSIEIKATKDIFHSKTMFDLYANIND 1573
PR	06-SEP-2001; 2001US-00948933.		QY	241 KPDNLTKMESGKLYSDKQYLKONILRKGYALKVTTNPKGKTMFDLYANIND 300
PR	25-OCT-2001; 2001US-0342923P.		QY	181 IVDGLAFAGDMRLFVKONDQKAEIKRMPKEKIKTSEPKVYVSSYGNVIELGEGLISK 240
PR	08-FEB-2002; 2002US-00072851.		DB	1574 KPDNLTKMESGKLYSDKQYLKONILRKGYALKVTTNPKGKTMFDLYANIND 1633
PR	06-MAR-2002; 2002US-0362699P.		QY	301 AKIQKANPLRASETTYIADSRNVEDGRSTOSVIMSLALGENITRYQVFPMKNDGEA 360
PA	(BLT-) BLTTRA PHARM INC.		DB	1634 AKIQKANPLRASETTYIADSRNVEDGRSTOSVIMSLALGENITRYQVFPMKNDGEA 1693

QV	361	IDKGQNLVTDSSKVLVLFGKDDKEYTGEDKFVNVERAKEDGSMILPFDTPKVNLSMDKNYFNP	420
Db	1694	IDKGQNLVTDSSKVLVLFGKDDKEYTGEDKFVNVERAKEDGSMILPFDTPKVNLSMDKNYFNP	1753
QY	421	SKSNKTYVRNPEPYLRGKISDKGGFWELRVNESVDNYLIGDLHDNTDPNKLNVK	480
Db	1754	SKSNKTYVRNPEPYLRGKISDKGGFWELRVNESVDNYLIGDLHDNTDPNKLNVK	1813
QY	481	DGDIMDGWIKDYKANGFPDKVTMDGNYLQGYSIDLNAKAVGVHYQFLYDNKPEVNID	540
Db	1814	DGDIMDGWIKDYKANGFPDKVTMDGNYLQGYSIDLNAKAVGVHYQFLYDNKPEVNID	1873
QY	541	PKGNTSIEYADGKSVFVNINDKRNNGFGEIQOHIYINGKEYTSNDIKOIDKLNIK	600
Db	1874	PKGNTSIEYADGKSVFVNINDKRNNGFGEIQOHIYINGKEYTSNDIKOIDKLNIK	1933
QY	601	IWKDFARNTTKEFILNKDGEVSELKPHRTVTIONGKEMSTVSEEDPILPVYKE	660
Db	1934	IWKDFARNTTKEFILNKDGEVSELKPHRTVTIONGKEMSTVSEEDPILPVYKE	1993
QY	661	LEKGYQFGWEISGFECKDAGYVINLSDKTIPKPFKKIEERKEENKPTDVKKDN	720
Db	1994	LEKGYQFGWEISGFECKDAGYVINLSDKTIPKPFKKIEERKEENKPTDVKKDN	2053
QY	721	PQVNHSQLNEHRSRKEDLOREHEHSQSDSTKDVTATVLDKONNISSTNNPK	773
Db	2054	PQVNHSQLNEHRSRKEDLOREHEHSQSDSTKDVTATVLDKONNISSTNNPK	2106
RESULT 4			
ADM92113			
ID	ADM92113	standard; protein; 2140 AA.	
AC	ADM92113;		
XX			
DT	03 -JUN-2004	(first entry)	
DE	S pneumoniae antigenic protein sequence SeqId310.		
KW	antibacterial; gene therapy; Streptococcus pneumoniae infection;		
KW	antigenic.		
OS	Streptococcus pneumoniae.		
XX			
PN	W02004020609-A2.		
XX			
PD	11-MAR-2004.		
PP	02-SBP-2003; 2003WO-US027401.		
PR	30-AUG-2002; 2002US-0407082P.		
XX			
PA	(TUFT ) UNIV TUFIS.		
PI	Camilli A, Hava DL;		
XX			
DR	WPI ; 2004-239189/22.		
DR	N-PSDB; ADM91876.		
XX			
PT	New Streptococcus pneumoniae nucleic acid molecules, useful for diagnosing, treating and preventing active infections of streptococcus pneumoniae.		
PT	Claim 27; SEQ ID NO 310; 123pp; English.		
CC	This invention relates to novel isolated Streptococcus pneumoniae nucleic acid molecules and the antigenic polypeptides encoded by them. The invention may be useful for the production of compounds with an antibacterial activity or for gene therapy. The nucleic acid molecules, compositions and methods disclosed are useful for treating Streptococcus pneumoniae infection. The present sequence is that of an S pneumoniae protein of the invention.		
RESULT 5			
AY81710			
ID	AY81710	standard; protein; 2120 AA.	
XX			
AC	AY81710;		
XX			
DT	02-JUN-2000	(first entry)	
XX			
DE	Streptococcus pneumoniae protein sequence ID3.		
XX			
CC	Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS; bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism; kidney disease; diabetes; immunosuppressive disorder; otitis media;		
CC			

C 1 P 0

KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO20006738-A2.  
 XX  
 DD 10-FEB-2000.  
 XX  
 PF 27-JUL-1999; 99WO-GB002452.  
 XX  
 PR 27-JUL-1998; 98GB-00016336.  
 PR 19-MAR-1999; 99US-0125328P.  
 XX  
 PA (MICR-) MICROBIAL TECHNICS LTD.  
 XX  
 PT Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
 XX  
 DR WPI; 2000-195301/17.  
 XX  
 N-PSDB; AA291806.

PT Streptococcal proteins and polynucleotides useful for diagnosis, treatment and prophylaxis of bacterial infections.  
 PS Claim 2, Page 41-42; 76pp; English.

This sequence represents a Streptococcus pneumoniae protein of the invention. The proteins (or their homologues, derivatives and/or fragments) are useful as immunogens or antigens. Immunogenic or antigenic compositions comprising the proteins are useful as vaccines and also in diagnostic assays. The sequences are useful for the detection or diagnosis of S. pneumoniae infection, by contacting a sample to be tested with them. Agents capable of antagonising, inhibiting or interfering with the function or expression of the protein or polypeptide are useful in medical compositions in the treatment or prophylaxis of S. pneumoniae infection. As the sequences can be used to treat S. pneumoniae infection, they can be used to treat bacterial pneumonia, which has high rates in young children, the elderly, and in patients with predisposing conditions such as asplenia, heart, lung and kidney disease, diabetes, alcoholism, or with immunosuppressive disorders, especially AIDS. They can also be used to treat pneumococcal septicemia, otitis media, sinusitis, and meningitis.

XX SQ Sequence 2120 AA;

Query Match Similarity 99.7%; Score 4015.5; DB 3; Length 2120;  
 Best Local Similarity 99.9%; Pred. No. 1-1e-223; Matches 773; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 KLGELIAEKKFKPKNGKKGSKLKDGTGVEHHQENESIKEKSFTIDRNISTRDENK 60  
 1313 KLGELIAEKKFKNGKKGSKLKDGTGVEHHQENESIKEKSFTIDRNISTRDENK 1372

QY 61 DLGKLKIKKFRETDFFSETGKMEEVYDVKYDDKGNTIAYDDCTDLEYETEKDEISKI 120  
 1373 DLGKLKIKKFRETDFFSETGKMEEVYDVKYDDKGNTIAYDDCTDLEYETEKDEISKI 1432

QY 121 YGVLSPSKDKGHFETLGKISNSVSKAKVYVGNNSKIETKATKDFHSKMTMDLYANIND 180  
 1433 YGVLSPSKDKGHEFILGKISNSVSKAKVYVGNNSKIETKATKDFHSKMTMDLYANIND 1492

QY 181 IVGLLAGFKAGDMRLFKVNDQKKAIEKIRMPKETKESPYVSSYGVNLGELEGDSK 240  
 1493 IVVGLLAGFKAGDMRLFKVNDQKKAIEKIRMPKETKESPYVSSYGVNLGELEGDSK 1552

Db 241 KPDNLTKQESGKYDSKQKVILKDNTLRKVALKTTPKGKTMLEMGVYSEDI 300  
 1553 KPDNLTKQESGKYDSKQKVILKDNTLRKVALKTTPKGKTMLEMGVYSEDI 1612

Db 301 AKIQOKANPLRALESETTYIADSARVEDGRSTOSVMSALDGENIIRQVFFPRNDIGEA 360  
 1613 AKIQOKANPLRALESETTYIADSARVEDGRSTOSVMSALDGENIIRQVFFPRNDIGEA 1672

Ps 361 IDKGNLVTDSKVLFKGDKDEKEYGEDGSMLPIDTKPVNLSDMDKNYPNP 420

Db 1673 IDKGNLVTDSKVLFKGDKDEKEYGEDGSMLPIDTKPVNLSDMDKNYPNP 1732  
 QY 421 SGSNKIVTRNPFPYRKSISDGKFNWELRINESVUDNYLYGDLHMDNTR-DENIKLNV 479  
 Db 1733 SGSNKIVTRNPFPYRKSISDGKFNWELRINESVUDNYLYGDLHMDNTR-DENIKLNV 1792

Db 1792 KODDIMGMKOKYANGPPDKUTDMGCVYLOTGYSDDLNKAVGVMQFLDNVKPEVNI 539  
 QY 480 KODDIMGMKOKYANGPPDKUTDMGCVYLOTGYSDDLNKAVGVMQFLDNVKPEVNI 1852

Db 1793 KODDIMGMKOKYANGPPDKUTDMGCVYLOTGYSDDLNKAVGVMQFLDNVKPEVNI 1852

QY 540 DPKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEQQIYINGKEYTSFRNDIKOIDKLN 599

Db 1853 DPKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEQQIYINGKEYTSFRNDIKOIDKLN 1912

Db 600 KIWKWDARNTTKEFLINKOTGEVSLPKPRVVTIQLNGKEMSTTVESEDFILPVTKG 659  
 QY 1912 KIWKWDARNTTKEFLINKOTGEVSLPKPRVVTIQLNGKEMSTTVESEDFILPVTKG 1972

Db 1972 ELEKGYQPDGWAISSGFGKKDAGYVINVLSKOTFIKPVKKLEBKCEBENKTDFVSKKD 719  
 Db 1973 ELEKGYQPDGWAISSGFGKKDAGYVINVLSKOTFIKPVKKLEBKCEBENKTDFVSKKD 2032

QY 720 NPQVNHSQLNESHHRKEDLQRERHSOKSDSTKDVATVLDKNNISSKSTTNHPNK 773  
 Db 2033 NPQVNHSQLNESHHRKEDLQREBHQSOKSDSTKDVATVLDKNNISSKSTTNHPNK 2086

Db 2086 RESULT 6 ADK48759  
 XX ADK48759 standard; protein; 2138 AA.  
 AC ADK48759;  
 XX DT 20-MAY-2004 (first entry)  
 XX DE Streptococcus pneumoniae protein, Seq ID No 5274.  
 XX KW Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.  
 OS Streptococcus pneumoniae.  
 XX PN US6693703-B1.  
 XX PD 02-MAR-2004.  
 XX PF 26-MAY-2000; 2000US-00583110.  
 PR 02-JUL-1997; 97US-0051533P.  
 PR 12-MAY-1998; 98US-0085131P.  
 PR 30-JUN-1998; 98US-00107433.

XX PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX DR Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Housewright CE;  
 XX DR WPI; 2004-212393/20.  
 N-PSDB; ADK46098.

XX PT New nucleic acid molecules and polypeptides useful for diagnosing, preventing and treating pathological conditions resulting from bacterial infection, e.g. Streptococcus pneumoniae infection, and in drug screening.

XX PS Disclosure; SEQ ID NO 5274; 301pp; English.

XX The invention relates to isolated Streptococcus pneumoniae nucleic acids and polypeptides. The nucleic acids and proteins are useful for diagnosing, preventing and treating pathological conditions resulting from bacterial infection, such as S. pneumoniae infection. These may also be used for drug screening procedures. The present sequence represents a Streptococcus pneumoniae polypeptide of the invention. Note: The sequence data for this patent did not appear in the printed specification but was



Db	255 KYYRNPFYRQKISKGCFNWLRTESWVNTLIVDIDHDTNPKINKVQSDI 314	CC
Qy	485 MDWGMKOKYKANGSPDKYPTMDGNYLQGSDNPKAVGWHYQFLYDNPKPEVNIDPKGN 544	CC
Db	315 MDWGMKOKYKANGSPDKYPTMDGNYLQGSDNPKAVGWHYQFLYDNPKPEVNIDPKGN 374	CC
Qy	545 TSIEYADGKSUVNFINDKRNGFDGETQOHITINGKWTSENDIKOIDKTNIKVK 604	CC
Db	375 TSIEYADGKSUVNFINDKRNGFDGETQOHITINGKWTSENDIKOIDKTNIKVK 434	CC
Qy	605 DPARNTTKEFLNKDGEVSELPKPHRTVTONGKEMSTIVSEDFILPVVKGELEKG 664	CC
Db	435 DPARNTTKEFLNKDGEVSELPKPHRTVTONGKEMSTIVSEDFILPVVKGELEKG 494	CC
Qy	665 YQFDGWEISLGFGSKKDGAYWVNLNSKDTIKPKVKKIEKEENPKPTDVKDDKNPQN 724	CC
Db	495 YQFDGWEISLGFGSKKDGAYWVNLNSKDTIKPKVKKIEKEENPKPTDVKDDKNPQN 554	CC
Qy	725 HSQLNESHRKEQLOREBSQKSSTDYATVLDANISSKSTINPK 773	CC
Db	555 HSQLNESHRKEQLOREBSQKSSTDYATVLDANISSKSTINPK 603	CC
Sequence 1529 AA;		
Query Match 18.6%; Score 750.5; DB 8; Length 1529;		
Best Local Similarity 79.7%; Pred. No. 1.3e-34;		
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;		
Qy	RESULT 8	SQ
ID	ADR9136	1 KLGEGAEESKENTGKESGKDKDGTGVENHQNEENESIKRSFIDTRNISTIRDPENK 60
XX	ADR9136 standard; protein; 1529 AA.	Db 1336 KLGEGAEESKENTGKESGKDKDGTGVENHQNEENESIKRSFIDTRNISTIRDPENK 120
AC	ADR9136;	Db 1396 DLKKUJKKERFVDDFTSETGRMMEYDVKDDGENIAYDGTDLIEYEPLKDEISKI 1454
DT	16-DEC-2004 (first entry)	Qy 121 YGVISPSKDKGHEIIGKISNVSKAKVYGNVYKIEKATKYDHSKTMFDLYANIND 180
DB	XX	Db 1455 YGVISPSKDKGHEIIGKISNVSKAKVYGNVYKIEKATKYDHSKTMFDLYANIND 1514
OS	Novel S. pneumoniae protein sequence, SEQ ID 4771.	Qy 181 YVDGLAF 187
KW	Menigitis; bacteraemia; pneumonia; otitis media; vaccine; bacterial infection.	Db 1515 YVDGLAF 1521
XX	Streptococcus pneumoniae.	RESUIT 9
XX	US6800744-B1.	AAW55096
XX	05-OCT-2004.	ID AAW55096 standard; protein; 117 AA.
PD	PP 30-JUN-1998; 98US-00107433.	XX AAW55096;
XX	PR 02-JUL-1997; 97US-0051553P.	DT 02-OCT-1998 (first entry)
PR	12-MAY-1998; 98US-0085131P.	XX DE Streptococcus pneumoniae SP0043 protein.
XX	PA (GENO-) GENOME THERAPEUTICS CORP.	XX KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis; detection; pneumonia; otitis media; meningitis.
XX	PI Doucette-Stamm LA, Bush D;	XX OS Streptococcus pneumoniae.
XX	WPI; 2004-697205/68.	PN WO9818930-A2.
DR	N-PSDB; ADR93533.	XX PD 07-MAY-1998.
XX	WPI; 2004-697205/68.	PP 30-OCT-1997; 97WO-US019422.
DR	N-PSDB; ADR93533.	PR 31-OCT-1996; 96US-0029960P.
XX	New isolated nucleic acid encoding a Streptococcus pneumoniae polypeptide, useful for diagnosing, preventing and/or treating pathological conditions resulting from the bacterial infection.	XX PA (HUMA-) HUMAN GENOME SCI INC.
XX	Disclosure; SEQ ID NO 4771; 151pp; English.	XX PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
CC	The invention relates to an isolated nucleic acid comprising a sequence encoding a Streptococcus pneumoniae ADR9136 polypeptide, or its fragments, with any of fully defined sequences (appearing as ADR94308, ADR94489, ADR94800, ADR94837, ADR94969, ADR95233, ADR95642, ADR95682, ADR91885, ADR91917, ADR92234, ADR93019, ADR93079, ADR92366, ADR92650 or ADR93776 or at least 20 or 30 consecutive nucleotides of the nucleotide sequences, or at least 40, 60 or 300 consecutive nucleotides, which is hybridisable under high stringency conditions to the nucleotide sequence. The nucleic acids and proteins are chosen from 5206 disclosed sequences. Also included are a recombinant expression vector comprising the isolated nucleic acid cited above operably linked to a transcription regulatory	XX DR WPI; 1998-272224/24.
CC		CC N-PSDB; AAV27357.
CC		PT Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae - or their epitope-containing fragments, useful in protective or therapeutic vaccines, and for diagnosis.
CC		PT therapeutic vaccines, and for diagnosis.
CC		PS Claim 11; Page 62; 118pp; English.
CC		CC The present sequence represents a protein from Streptococcus pneumoniae protein.
CC		CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein

CC can be useful in vaccines for inducing protective antibodies against  
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
 CC are used to detect Streptococcus infection (by usual hybridisation or  
 CC amplification methods), also for isolating Streptococcus genes or their  
 CC allelic variants. The protein can be used similarly to detect specific  
 CC antibodies in standard immunoassays, especially for diagnosing or  
 CC monitoring infections. Antibodies which bind the protein are used to  
 CC detect corresponding antigens, to purify the protein and for passive  
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
 CC (especially 10-300) µg/ml per dose  
 XX sequence 117 AA;  
 SQ Query Match 15.3%; Score 615; DB 2; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-28;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 657 YKGELEKGYQPGWEISGPFEGKDKAGYVINLISKDFIKPVKKIEKEENKPTFDVK 716  
 Db 1 YKGELEKGYQPGWEISGPFEGKDKAGYVINLISKDFIKPVKKIEKEENKPTFDVK 60  
 QY 717 KKDNPQVNHSQLNESHRKEDLQREERHQSKSISTKDVATVLDKNNISSKSTTNNPK 773  
 Db 61 KKDNPQVNHSQLNESHRKEDLQREERHQSKSISTKDVATVLDKNNISSKSTTNNPK 117

RESULT 10  
 ABB54590  
 ID ABB54590 standard; protein; 117 AA.  
 AC ABB54590;  
 XX DT 04-SEP-2002 (first entry)  
 XX DE S. pneumoniae SP043 protein sequence SEQ ID NO:68.  
 XX KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
 XX OS Streptococcus pneumoniae.  
 XX PR US2002061545-A1.  
 XX PD 23-MAY-2002.  
 XX PR 22-JAN-2001; 2001US-00765272.  
 PR 30-OCT-1997; 97US-00961083.  
 XX PA (CHOI) / CHOI G H.  
 PA (KUNS) / KUNSCHE CA.  
 PA (BARA) / BARASH S C.  
 PA (DILL) / DILLON P J.  
 PA (DOUG) / DOUGHERTY B.  
 PA (FANN) / FANNON M R.  
 PA (ROSE) / ROSEN C A.  
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 PI Rosen CA;  
 XX DR WPI; 2003-764574/72.  
 DR N-PSDB; ADC45148.  
 XX PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides  
 PT useful for producing vaccines for prevention or attenuation of infection  
 PT by Streptococcus pneumoniae.  
 XX PS Example 1; SEQ ID NO 68; 58pp; English.  
 XX The invention relates to an isolated polynucleotide consisting of a  
 CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding  
 CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae  
 CC antigens. Also included are making a recombinant vector by inserting the  
 CC nucleic acid into a vector, an isolated polynucleotide consisting of at  
 least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a  
 CC recombinant host cell comprising the SP028 polynucleotide. The nucleic  
 CC acids are useful as DNA vaccine against Streptococcus pneumoniae  
 CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae  
 CC antigen nucleic acids are useful as probes for use in diagnostic methods  
 CC for detecting S. pneumoniae gene expression. The present sequence  
 CC represents an S. pneumoniae antigenic protein.

CC pneumoniae antigens have antibacterial activity and can be used in  
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or  
 CC attenuate a Streptococcal infection in an animal. The polynucleotides  
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus  
 CC nucleic acids. ABQ4905 to ABQ4910 represent primers used in the cloning  
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example  
 CC from the present invention.

Query Match 15.3%; Score 615; DB 5; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-28;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 657 YKGELEKGYQPGWEISGPFEGKDKAGYVINLISKDFIKPVKKIEKEENKPTFDVK 716  
 Db 1 YKGELEKGYQPGWEISGPFEGKDKAGYVINLISKDFIKPVKKIEKEENKPTFDVK 60  
 QY 717 KKDNPQVNHSQLNESHRKEDLQREERHQSKSISTKDVATVLDKNNISSKSTTNNPK 773  
 Db 61 KKDNPQVNHSQLNESHRKEDLQREERHQSKSISTKDVATVLDKNNISSKSTTNNPK 117

RESULT 11  
 ADC45149  
 ID ADC45149 standard; protein; 117 AA.  
 AC ADC45149;  
 XX DT 18-DEC-2003 (first entry)  
 XX DE S. pneumoniae antigenic protein SP043.  
 XX DR Antigen; bacterial infection; vaccine; pneumonia; antibacterial.  
 XX KW Streptococcus pneumoniae.  
 XX OS Streptococcus pneumoniae.  
 XX PN US6573082-B1.  
 XX PR 30-OCT-1996; 96US-0029960P.  
 XX DD 03-JUN-2003.  
 XX PR 28-MAR-2000; 2000US-00536784.  
 XX PN 31-OCT-1996; 96US-0029960P.  
 XX PR 30-OCT-1997; 97US-00961083.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PT Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 PT Rosen CA;  
 XX DR WPI; 2003-764574/72.  
 DR N-PSDB; ADC45148.  
 XX PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides  
 PT useful for producing vaccines for prevention or attenuation of infection  
 PT by Streptococcus pneumoniae.  
 XX PS Example 1; SEQ ID NO 68; 58pp; English.  
 XX The invention relates to an isolated polynucleotide consisting of a  
 CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding  
 CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae  
 CC antigens. Also included are making a recombinant vector by inserting the  
 CC nucleic acid into a vector, an isolated polynucleotide consisting of at  
 least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a  
 CC recombinant host cell comprising the SP028 polynucleotide. The nucleic  
 CC acids are useful as DNA vaccine against Streptococcus pneumoniae  
 CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae  
 CC antigen nucleic acids are useful as probes for use in diagnostic methods  
 CC for detecting S. pneumoniae gene expression. The present sequence  
 CC represents an S. pneumoniae antigenic protein.

ABQ8492 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.

SQ	Sequence 117 AA.	Query Match	Score 15-34;	DB 7;	Length 117;
		Best Local Similarity	100.0%;	Pred. No.	3.2e-28;
		Matches	117;	Conservative	0;
		Mismatches	0;	Indels	0;
		Gaps	0		
QY	657 YKGEIIRKGIQFDGWESISGPEGKDDAGVWNISSKDTFKPVFKKIREKKEEENKFPTDYSK 716				
Db	1 YKGEIIRKGIQFDGWESISGPEGKDDAGVWNISSKDTFKPVFKKIREKKEEENKFPTDYSK 60				
QY	717 KKDNFQVNHSQLNEESEHKKEDLQREHRSOKSPSTDKVATVLDKNNISSKTTNNPNK 773				
Db	61 KKDNFQVNHSQLNEESEHKKEDLQREHRSOKSPSTDKVATVLDKNNISSKTTNNPNK 117				
RESULT 12					
ABU24813	ABU24813 standard; protein; 1196 AA.				
XX	AC ABU24813;				
XX	DT 19-JUN-2003 (first entry)				
DB	DE Protein encoded by Prokaryotic essential gene #10340.				
XX	KW Antisense; prokaryotic essential gene; cell proliferation; drug design.				
XX	OS Clostridium botulinum.				
XX	CLOSTRIDIUM BOTULINUM.				
XX	PN WO200277183-A2.				
XX	PD 03-OCT-2002.				
XX	PP 21-MAR-2002; 2002WO-US009107.				
XX	PR 21-MAR-2001; 2001US-00815242.				
PR	06-SEP-2001; 2001US-00948993.				
PR	25-OCT-2001; 2001US-0342923P.				
PR	08-FEB-2002; 2002US-034292851.				
PR	06-MAR-2002; 2002US-0362699P.				
XX	PA (ELIT-) BLITRA PHARM INC.				
XX	PI Wang L, Zamudio C, Malone C, Harelbeck R, Ohlsen KL, Zyskind JW;				
PI	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;				
XX	DR WII; 2003-029926/02.				
DR	N-PSDB; ACA28683.				
XX	PT New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for potential drug discovery programs.				
XX	PT Claim 25; SEQ ID NO 52737; 1766pp; English.				
CC	The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification, where expression of the nucleic acid inhibits proliferation of a cell. Also included are:				
CC	(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene				

Db 729 EL---VYNNELITKEKGKLNAAKEDTERIINSSINSRERIKUTOKIKDENTINESQK 784  
 QY 710 PTFDVSKKKDNPIQVNHSQNLNEHSHRKEQDQRERSQKSSTK---DVTATVLDKONISSK 765  
 Db 785 QIEELKUKKD--LNHENDIKEC--EDTLQNESENVKQIKDKQLEYKIEKAQAKLDEMUSIK 839  
 QY 766 -----SPTNNPK 773  
 Db 840 KELYSMDTNITNLNNENK 857

RESULT 13

ABO23608 standard; protein; 861 AA.  
 XX  
 AC ABO23608;  
 XX  
 DT 04-SEP-2003 (first entry)

DE Plasmodium falciparum outlier protein #5.

KW Candidate protein identification; pathogen; anti-infective; outlier protein; virulence protein; antigen; drug target protein; pathogenic organism; antimicrobial.

OS Plasmodium falciparum.

XX US2003039963-A1.

XX PD 27-FEB-2003.

XX 30-MAR-2001; 2001US-00820843.

PR 30-MAR-2001; 2001US-00820843.

XX PA (BRAH/)  
 (RAMA/)  
 (NAND/)  
 (BHIM/)  
 BRAHMACHARI S K.  
 RAMACHANDRAN S.  
 Nandi T.  
 Bhimrao C.

XX PI Brahmachari SK, Ramachandran S, Nandi T, Bhimrao C;

DR WPI; 2003-492159/46.

XX PT Identifying candidate proteins useful as anti-infectives involves matching outlier protein sequences with protein sequences in databases.

XX PS Example 7; Page 95-97, 117pp; English.

CC The present invention relates to a method for identifying candidate proteins in pathogens useful as anti-infectives. The invention discloses a computational method which involves the calculation of several sequence attributes and their subsequence analysis results in the identification of outlier proteins in different pathogens. The method is useful for the identification of outlier proteins (e.g. virulence proteins, antigens or proteins used as drug targets) in pathogenic organisms. The method of the invention provides reproducible results as it does not depend on the variable biochemical characterisation of proteins. ABO23500-ABO23617 represent outlier proteins identified from different pathogenic organisms

SQ Sequence 861 AA;

Query Match 5.5%; Score 222; DB 7; Length 861;  
 Best Local Similarity 19.3%; Pred. No. 0.00026;  
 Matches 162; Conservative 136; Mismatches 324; Indels 216; Gaps 37;

QY 2 LGEIAESKFKNGG-----KEGSILKOTGVEHH--QENEESIKEKSSFTI 47  
 Db 154 IANNINGKVKNUDDCNVLPTNLYIDKEG--KMHJTGKEHYAASSNEYHNHNNTNNY 211  
 QY 48 DMRNISTIRDPEKNDL-----KKLUKKEPREVDDFTSFTGKRMEYDYKDDGKII 98

Db 212 NNNSYNNNNFCNNNNNNNNNNNNNNNNKGMGNKVERSINYLK--KEHDMDYEYNNKGNR 268  
 QY 99 AYD----DGTDLFETEKLDEIKSKYGVLSPLSPKGFFET--LGKISVNSKNAKYGN 151  
 Db 269 KNDSEKWTMDNPFLHY-----SKGNVDIFTGLDIKKVAKONEGGN 310  
 QY 152 NYKSIKIKATQDFKSTKMTFDLYANINDIVDG---LAFAQADMRLFVKONDOKKABKIR 208  
 Db 311 KYMM----HDNNSN-----NSNNVANNNNNSNSNSNNNNIFKONDENI---- 351  
 QY 209 MPEKIKETKSYPYVSSYGNVIELGG-DISK--NKPDNLTKMSGKSYSDSEKQQYL 264  
 Db 352 -----TKSINPAKWFRRNNNNNNVNETTDIYLNK-----NSQGHSDGRNNNN 396  
 QY 265 KUNIILRKGVALKVTTYNGPKTDLLEGNGYTSKEDIAKIQKANPN--LRALESETTIYAD 321  
 Db 397 GNNTI-----NNNNNNNNNIFGNNSNTENYNTNNNNNNNNNNNNNNNNNNNNNN 447  
 Db 322 SRNVEDGRSTOSVLMALSDGFNIIRYQVFTPKANQK--EAIKDQDNLVTDSSKLVLPGD 380  
 QY 448 NINTNSGREGEEKISINTVAE--LIMKQISWIKERNKGKLDVLEKKT-----FGFL 494  
 Db 495 DNYYQNGSNN-----NSLEKNNM---KENDIYSKESASKRIMDIFR 533  
 QY 436 ---RSKISDKXGGFWELRVRNESVUDMLTYGDHLIDNTDFNPKLNVRQGDIMDWGMD 491  
 Db 534 TLNSNGLVPSQ-----SLLVNOVSUANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 582  
 QY 492 YKANGPFDKVTDMDGAVYVQYLOGYSIDLAK--AVGHYQOLYDNKPEVWLPKNTIE 548  
 Db 583 -----NN 633  
 PR 549 YADGKSVVFNTNDKRGFGPCBEOFOHIIYINGKEYTSFNDIKQIIDK----- 595  
 Db 634 NSNNVVEHYIYNNNNK--PKRKNINHLPDKNNNNNNNNNNNNNNNNNNNNNNNNNN 691  
 QY 596 TLNLIKVKOFARN--TTWKEFILNKDGPGEVSELKPHRTVTONGKEMSTIVSEED 651  
 Db 692 SFNSNDINNNKONAQNINNTIPINSLRLD-NEVDVHNNSNSENTONAK--VSNTLDSL 748  
 QY 652 FILPVKGELKEKGQDFGWEISGFGCKDAGVYVLSKDTFR-----PVFKIEKK 704  
 Db 749 SILKASKSQGANNYNT--KNEFNNNN-----NNNNSFINTNSQVQYPSHQQQQOQH 799  
 QY 705 EBENKPFDPVSKKKQVHRSQNLNBRSRKEDLBREBSOK--SDSTKQVATVLDK 759  
 Ps 800 QQQQQ---QQQQQLIQTQINSTHUNDPKKFNKERRYKPYFEDGTINETMTRK 855  
 Db

RESULT 14

ID ADK99186 standard; protein; 1166 AA.

XX AC ADK99186;  
 XX DT 20-MAY-2004 (first entry)

XX Streptococcus agalactiae ORF SAG0416-related protein 11.

XX immunogenic composition; group B Streptococcus; GBS; antibacterial; streptococcal infection; vaccine; SAG.

XX OS Streptococcus agalactiae.

XX PN WO2004018646-A2.

XX DD 04-MAR-2004.

XX PR 26-AUG-2003; 2003WO-US026827.

XX 26-AUG-2002; 2002US-0406237P.



CC the native pattern of folding. Larger amounts of the protein can be  
 CC produced recombinantly than would be possible using the parasites as  
 CC source. (Updated on 25-MAR-2003 to correct PR field.)

XX Sequence 1639 AA;

Query	Match	Score	DB	Length	Best Local Similarity	Local Similarity	Pred.	No.	Mismatches	Matches	Conservative	Indels	Gaps	39;	
Qy	23	KDTG-VERHROENERSIK-----KSSFTIDRNISTIRPFENKOLKKLKKKFREVD	5.4%	218;	DB 2;	Length 1639;									
Db	232	KDNVGKMDYTKRKKNKTIENTNELTBESKKTIDKNNKATRKE-----KKCQYQAOY	18.6%	0 0011;	Pred.	No. 0 0011;									
Qy	76	FTSETGKRMEEVDYKDKGKIIAYDGTIDYEYETEKLDEKSKYIVGLVLSPSKGHPEIL	75												
Db	284	DISIYVKOLEE-----AHNLISV----LE--KRIDTK-----KNEKELL	135												
Qy	136	GKTSNVSKNAKYGGNNYKSEIKATKTYDFI-----SKMTFOLYANTINDIVGGLAFA	188												
Db	320	DKINEIKNPPPANSGNTPTNLLDKNNKKEEHEKEIAKTIKTFNIDSFLFDPL-----	373												
Qy	189	GIMRLFVKDNOKKAETKIRNPKEKEKTSFSP-----YVSYGNV-----TEIGE-GDL	239												
Db	374	EBLEVYLREKN-KNNDISAKVSTKESTEPNYPNGVTPLYPLSYNDINNALNEINSFGDLI-	430												
Qy	240	NKPDNLTKMEEGKIXYDSEKKQYL-LKDNLIL-----RKGIALKVTYTPGKTDM	289												
Db	431	-NPFDYTKERPSKNIYTDRNKKFINEKIEKKEKIEKESDKSYDRSKLN---DIT	485												
Qy	290	EG-----NGVSE-----DIAKIOKA-----NPNALSETTIYADSARNEVEDRGSQV	334												
Db	486	KEYEKULNETVDSKFNNDLTNFECOMGKSYKUEKLTHTFASYENSKHNLBKLTK	545												
Qy	335	IMSALDFR-----NII-----RTQVFPRKMNDGEAIDKGHLVUTDSKLV-FGKODKEYT	385												
Db	546	AUKYMEIDLSELRNIVVEKELVYKNUISKLEMIELTVE- -NIKODEQLP-----ERKIT	598												
Qy	386	GRDKFKNVAEAKEDGSMFLFDPKPVNLSMDKQYFPKSNKIVRNFBPYRGKISKPGF	445												
Db	599	IKDENKDEKILKLEVDSDIVKVQVOQVULL--MNKIDELEKKTQJLTKNTE-LHNIHVPN	653												
Qy	446	WELRNVESVVDVNLIGDHAIDNTDPMKL-----ANV-----	480												
Db	654	KOENKQB-----PYLILWIKKEKDQLKVEMPKVSELINEEKNIKITEGQSDDNSEPSTEGL	709												
Qy	481	-----DGDIMDWGMKDYK-----ANGFPDKV	501												
Db	710	TGQATKPGQGQAGSALEGDSVQAQAEQKQAOQPPVPUVPUVPSAKAQVTPPPAPVNUK	769												
Qy	502	TDMGNYTLQTCYSDMA-----KAVGVHYOPLYDNKPENIDPKGNTSEYADCKSVV	556												
Db	770	SKLD-----YLEKLYEFMLVSYICHKYTLVSHSTMNEKILKQKTKTKEBKESKUSCDPL	826												
Qy	557	FRIN-----DKRNGFEG-----EIQEPMIYINGKEYTSFNDIKQIQLINKIV	603												
Db	827	FNIIONNIPVMMSMFDSLNSISQLEMEIYKEMWCMLYKLKNDKIKNLPEAKVYSTV	886												
Qy	604	KDFARNITVKEFILNKDGTGEVSELKPHRVVTIIONGKEMSTIVSBDFFLPVYGELEK	663												
Db	887	KLISSSSQWPLSLTPQDKPESANDTSHSTNLANSKLFNPLS-----LGKNN	937												
Qy	664	GYOPDGWEISGPSEGK-----KDAFYVILSKDPTPKPFKKERKEFENKPTP-D	713												
Db	938	IYQ-----ELIGOKSENFPYEKLIKDSDFYHESFTNEVFKSADDINSLNDSSKRKLEED	993												
Qy	714	VSKKKKQPVQHSQLESHRSKEDLQREHSSQXSDSKDVTATLDRONISK-STINNP	771												
Db	994	INKKKTQIQLSFIDLYNKYKLKLERLFDKTTVQGKVKYKNOIKUTLKEQLESKLN	1052												



Db 1632 AKIQKANPILRALKSETTIYADSRNVEDGRSTQAVIMSLDFNIRYQFTPKMDKGEA 1691  
 Qy 361 IDKDGNIUTDSKUVLFGKDKDKEYTGEDKFNEAIKEGDSMLFDTKPVNLNSMDKRYFNP 420  
 Db 1692 IDKDGNIUTDSKLVLFGKDKDKEYTGEDKSNEAIKEGDSMLFDTKPVNLNSMDKRYFNP 1751  
 Qy 421 SGSNKTYRNPEFYLGKISDKGGFENELRNESVVDNLILYGLDHLIDTRDPNKLVK 480  
 Db 1752 SKSNKDYRNPEFYLGKISDKGGFENELRNESVVDNLILYGLDHLIDTRDPNKLVK 1811  
 Qy 481 DGDIMDWMGKOKXKANGRPDKYTDMDGNVYLQGYSUNAKVGHVQFLDNVKPBNID 540  
 Db 1812 DGDIMDWMGKOKXKANGRPDKYTDMDGNVYLQGYSUNAKVGHVQFLDNVKPBNID 1871  
 Qy 1872 PKGNTSLEYAQQSVVFNINDKRNNFGDGETQEOHYINGKEYTSNDIKOIKDTLNK 1931  
 Qy 601 IWKDFAARTTKEFLNKDGESELKPHRTVTIONKEMSTIVSEEDFILPVYKE 660  
 Db 1932 IWKDFAARTTKEFLNKDGESELKPHRTVTIONKEMSTIVSEEDFILPVYKE 1991  
 Qy 661 IERKYOPDGWELISGPGKKGDDAGYVNLSKOTFIKPVKKEBEKENKPTFDVSKKDN 720  
 Db 1992 IERKYOPDGWELISGPGKKGDDAGYVNLSKOTFIKPVKKEBEKENKPTFDVSKKDN 2051  
 Qy 721 PQVNHSQLNESHRKELOREHSQKSDSTKVATVLDKNNISSKTTNPNK 773  
 Db 2052 PQVNHSQLNESHRKELOREHSQKSDSTKVATVLDKNNISSKTTNPNK 2104

RESULT 2  
 US-05-107-433-3169  
 ; Sequence 3169, Application US/09107433  
 ; Patent No. 6800744  
 ; GENERAL INFORMATION:  
 APPLICANT: Lynn A Doucette-Stamm and David Bush  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
 SEQUENCES RELATING TO STREPTOCOCUS PNEUMONIAE FOR DIAGNOSTIC  
 THERAPEUTICS

NUMBER OF SEQUENCES: 5206  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02354

COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD/ROM ISO9660  
 COMPUTER: <Unknown>  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: <Unknown>

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107-433  
 FILING DATE: 30-Jun-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/ 095131  
 FILING DATE: May 12, 1998  
 APPLICATION NUMBER: 60/051553  
 FILING DATE: July 2, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Arinello, Pamela Deneke  
 REGISTRATION NUMBER: 40,489  
 REFERENCE DOCKET NUMBER: GTC-011  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781)893-5007  
 TELEFAX: (781)893-8277  
 INFORMATION FOR SEQ ID NO: 3169:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 637 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear

RESULT 2  
 US-05-107-433-3169  
 ; Sequence 3169, Application US/09107433  
 ; Patent No. 6800744  
 ; GENERAL INFORMATION:  
 APPLICANT: Lynn A Doucette-Stamm and David Bush  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
 SEQUENCES RELATING TO STREPTOCOCUS PNEUMONIAE FOR DIAGNOSTIC  
 THERAPEUTICS

NUMBER OF SEQUENCES: 5206  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02354

COMPUTER READABLE FORM:

MOLECULE TYPE: protein.  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Streptococcus pneumoniae  
 FEATURE:  
 NAME/KEY: misc. feature  
 LOCATION: (B) LOCATION 1.. 637  
 SEQUENCE DESCRIPTION: SEQ ID NO: 3169:  
 ; US-05-107-433-3169  
 ; Query Match 75.7%; Score 3048; DB 4; Length 637;  
 Best Local Similarity 99.3%; Pred. No. 5.6e-19; 1; Indels 0; Gaps 0;  
 Matches 585; Conservative 3; Mismatches 1;  
 Query 185 LAPAGDMRLFKNDOKKAETKIRKPEKIKETSKSYPPYSQGNVIELEGQDLISKPNP 244  
 Db 15 LAFAGDMRLFKNDOKKAETKIRKPEKIKETSKSYPPYSQGNVIELEGQDLISKPNP 74  
 Qy 245 LTKMSESGKTSDESKOQYQJLKDNTIRKGALKVUTNPGKTDMLEGNGYSKEDAKIQ 304  
 Db 75 LTKMSESGKTSDESKOQYQJLKDNTIRKGALKVUTNPGKTDMLEGNGYSKEDAKIQ 134  
 Qy 305 KANPNRALARBTIYADSRNVEDGSTQSUTMSALDGENITYQFTFKNDKRAIDK 364  
 Db 135 KANPNRALARBTIYADSRNVEDGSTQSUTMSALDGENITYQFTFKNDKRAIDK 194  
 Qy 365 GNLVTDSSKUVLFGKDKDKEYTGEDKSNEAIKEGDSMLFDTKPVNLNSMDKRYFNP 424  
 Db 195 GNLVTDSSKUVLFGKDKDKEYTGEDKSNEAIKEGDSMLFDTKPVNLNSMDKRYFNP 254  
 Qy 425 KLYVRNPEFYLGKISDKGGFENELRNESVVDNLILYGLDHLIDTRDPNKLVKD 484  
 Db 255 KLYVRNPEFYLGKISDKGGFENELRNESVVDNLILYGLDHLIDTRDPNKLVKD 314  
 Qy 485 MDWGMKDYKANGPDKYTDMDGNVYLQTGTSDSLNAKAVGPHYQFLDNVYKEPVNIDPKGN 544  
 Db 315 MDWGMKDYKANGPDKYTDMDGNVYLQTGTSDSLNAKAVGPHYQFLDNVYKEPVNIDPKGN 374  
 Qy 545 TSIYEYGKSVVFNINDKRNGFDEBIQEOHYINGKEYTSNDIKOIKDTLNKIVK 604  
 Db 375 TSIYEYGKSVVFNINDKRNGFDEBIQEOHYINGKEYTSNDIKOIKDTLNKIVK 434  
 Qy 605 DEARNTTKEFLNKDGESELKPHRTVTIONKEMSTIVSEEDFILPVYKELEK 664  
 Db 435 DFARNTTKEFLNKDGESELKPHRTVTIONKEMSTIVSEEDFILPVYKELEK 494  
 Qy 665 YQFDGWEIISGPGKKGDDAGYVNLSKOTFIKPVKKEBEKENKPTFDVSKKDNPQN 724  
 Db 495 YQFDGWEIISGPGKKGDDAGYVNLSKOTFIKPVKKEBEKENKPTFDVSKKDNPQN 554  
 Qy 725 HSQLNESHRKELOREHSQKSDSTKVATVLDKNNISSKTTNPNK 773  
 Db 555 HSQLNESHRKELOREHSQKSDSTKVATVLDKNNISSKTTNPNK 603

RESULT 3  
 US-05-107-433-4771  
 ; Sequence 4771, Application US/09107433  
 ; Patent No. 6800744  
 ; GENERAL INFORMATION:  
 APPLICANT: Lynn A Doucette-Stamm and David Bush  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
 SEQUENCES RELATING TO STREPTOCOCUS PNEUMONIAE FOR DIAGNOSTIC  
 THERAPEUTICS

NUMBER OF SEQUENCES: 5206  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02354

MEDIUM TYPE: CDD/ROM ISO9660  
 COMPUTER: <Unknown>  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: <Unknown>  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107,433  
 FILING DATE: 30-Jun-1998  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 60 / 085131  
 FILING DATE: May 12, 1998  
 APPLICATION NUMBER: 60/051553  
 REFERENCE/DOCKET NUMBER: GTC-011  
 TELECOMMUNICATION INFORMATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Arniello, Pamela Deneke  
 REGISTRATION NUMBER: 40,489  
 REFERENCE/DOCKET NUMBER: GTC-011  
 TELEPHONE: (781) 893-5007  
 TELEFAX: (781) 893-9277  
 INFORMATION FOR SEQ ID NO: 4771:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1529 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Streptococcus pneumoniae  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (B) LOCATION 1...1529  
 SEQUENCE DESCRIPTION: SEQ ID NO: 4771:  
 ;  
 US-09-107-433-4771  
 Query Match 18.6%; Score 750.5; DB 4; Length 1529;  
 Best Local Similarity 79.7%; Pred. No. 1.8e+40; Mismatches 23; Indels 1; Gaps 1;  
 Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;  
 Qy 1 KUGRIAESKEPKNGLENGKGEGLSKKOTGTGVEHHQENNEESTEKSSTIDRNISTIDPFENK 60  
 Db 1336 KUGBIPESKEPKNKLKVKKDSDLNKETAEVENNLVUNOSTEGKSLNHIKHTISTRDPENK 1395  
 Qy 61 DLKLUKIKKPFREVDPFSETGKRMVEYDYKDDKKNIAVDDGPDLEYETEKLUDBIISKI 120  
 Db 1396 DLKLUKIKKPFREVDPFSETGKRMVEYDYKDDKKNIAVDDGPDLEYETEKLUDBIISKI 120  
 Qy 121 YGVISLSPSKOQHFELIGKISVSKAVVYGNNSIEIKATKYDHISKTMPFLANIND 180  
 Db 1455 YGVISLSPSKOQHFELIGKISVSKAVVYGNNSIEIKATKYDHISKTMPFLANIND 1514  
 Qy 181 IVDGGLAP 187  
 Db 1515 IVDGISP 1521  
 ;  
 RESULT 4  
 US-08-961-083-68  
 Sequence 68, Application US/08961083  
 Patient No. 6152469  
 GENERAL INFORMATION:  
 APPLICANT: Choi et. al.  
 TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
 NUMBER OF SEQUENCES: 452  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville  
 STATE: Maryland  
 COUNTRY: USA  
 ZIP: 20850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 COMPUTER: HP Vectra 486/33  
 OPERATING SYSTEM: MSDOS version 6.2  
 SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/536,784  
 FILING DATE: 30-Oct-1997  
 CLASSIFICATION: <Unknown>  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 08/961,083  
 FILING DATE: OCT-30-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Michelle S. Marks  
 REGISTRATION NUMBER: 41,971  
 REFERENCE/DOCKET NUMBER: PB340P3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (301) 309-8504  
 INFORMATION FOR SEQ ID NO: 68:  
 ;  
 US-08-961-083-68  
 Query Match 15.3%; Score 615; DB 3; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 3.9e+33; Mismatches 0; Indels 0; Gaps 0;  
 Matches 117; Conservative 0; Mismatches 0;  
 Qy 657 YKGEGLEKGYQFGWELSGFEGKRDAGVINTLSDFTRKPVPKIEKEGENKPTFDVK 716  
 Db 61 KCDNPQNHQSJNEBHRKDLOREBEHQSKS DSTKDVATVLDGNISKS TTNPNK 773  
 Qy 717 KCDNPQNHQSJNEBHRKDLOREBEHQSKS DSTKDVATVLDGNISKS TTNPNK 117  
 Db 61 KCDNPQNHQSJNEBHRKDLOREBEHQSKS DSTKDVATVLDGNISKS TTNPNK 60  
 ;  
 RESULT 5  
 US-09-536-784-68  
 Sequence 68, Application US/09536784  
 Patient No. 6573082  
 GENERAL INFORMATION:  
 APPLICANT: Choi et. al.  
 TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
 NUMBER OF SEQUENCES: 452  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville  
 STATE: Maryland  
 COUNTRY: USA  
 ZIP: 20850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 COMPUTER: HP Vectra 486/33  
 OPERATING SYSTEM: MSDOS version 6.2  
 SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/536,784  
 FILING DATE: 30-Oct-1997  
 CLASSIFICATION: <Unknown>  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 08/961,083  
 FILING DATE: OCT-30-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Michelle S. Marks  
 REGISTRATION NUMBER: 41,971  
 REFERENCE/DOCKET NUMBER: PB340P3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (301) 309-8504  
 INFORMATION FOR SEQ ID NO: 68:  
 ;

SEQUENCE CHARACTERISTICS:  
 LENGTH: 117 amino acids  
 TYPE: amino acid  
 STRANDBEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
 us-09-536-784-68

Query Match 15.3%; Score 615; DB 4; Length 117;  
 Best Local Similarity 100.0%; Pred No. 3.9e-33;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYOPDGWEISGFECKDAGTYVINSKDTF1PKPVKIEKEENKPTFDVSK 716  
 1 YKGELEKGYOPDGWEISGFECKDAGTYVINSKDTF1PKPVKIEKEENKPTFDVSK 60

QY 717 KKDNPQHNSQLNESRKEDIQREHEQSOKSISTKDTATIDKNNNKSSTNNPK 773  
 61 KKDNPQHNSQLNESRKEDIQREHEQSOKSISTKDTATIDKNNNKSSTNNPK 117

RESULT 6  
 Sequence 7, Application US/09200650E  
 General Information:  
 Applicant: Patti, Joseph M.  
 Applicant: Foster, Timothy J.  
 Applicant: Hock, Magnus A.O.  
 Applicant: Bidhim, Beidre Ni.  
 Applicant: Perkins, Samuel L.  
 Title of Invention: Extracellular Matrix Binding Proteins from Staphylococcus aureus  
 File Reference: P06283US2/BAS  
 Current Application Number: US/09/200,650B  
 Prior Filing Date: 1998-11-25  
 Prior Application Number: 60/066,815  
 Prior Filing Date: 1997-11-26  
 Prior Application Number: 60/098,427  
 Prior Filing Date: 1998-08-31  
 Number of SEQ ID NOS: 23  
 Software: Patentin Ver. 2.0  
 SEQ ID NO: 7  
 LENGTH: 1166  
 TYPE: PRT  
 ORGANISM: Staphylococcus aureus  
 us-09-200-650B-7

Query Match 5.0%; Score 203; DB 4; Length 1166;  
 Best Local Similarity 21.6%; Pred. No. 6.8e-05; Gaps 53;  
 Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;

QY 33 QNEEESTKE-KSSFTIDRNSTIRDPEPKNLKKLTKKKF--EVDDFTS 78  
 226 KNPKELKELYRNDSTWDHKSPKAPTAPEKAKRPNKAKRFAVOPAANVNDLIK 285

QY 79 EFGKRMEEYDKYDDKGNTIADGDTLEYTE-----KLDEBKSKYGVLS 125  
 286 VTKQTIR---VGDKDNVAAHGDQIEYDTEFTIDNKVKRGDTWINTYDRNVPSDLT 341

QY 126 PSKD-----GHEFLGKSNVSKAK---VYGNMYSIEKATKIVHFSK-----168  
 342 DKNDPDIRTDPSGEVIAKGTIDKATKQITYFTDVYKDEIKSRUTLYSDIKKTVPNE 401

Db 402 TSLNLTPTAKETGSQVTVYQDPWIK-GDSNISIFTKIDEDKOTIEQIYVPLKS 460

QY 169 ---TMTP-----DLYANIN-DIVDGIAFAGD---MRLFVK-DNDQKKAETKIRM-PEKIT 214  
 215 EFKSEY---PTVSSYGNVILGEQGD--LSKNKPDLTM-----ESGIYSESEKQO 261

Db 461 AANTKVDIAGSQVDDKCN-IKLNGNSTIDQTEIKVYKVNQDQLPQSNRiyDFSOYED 519

QY 262 YLLK-DNLKKGIAKVTTNPKGIDMLEANGVSKEDIAKIQKANPMLASETTIA 320

RESULT 7  
 Sequence 8, Application US/08973462B  
 General Information:  
 Applicant: Druilhe, Pierre  
 Applicant: Daubies, Pierre  
 Title of Invention: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
 File Reference: 0660-0125-0 PCT  
 Current Application Number: US/08/973,462B  
 Current Filing Date: 1998-02-05  
 Earlier Application Number: PCT/FR96/00894  
 Earlier Filing Date: 1996-06-12  
 Earlier Application Number: FR 95/07007  
 Earlier Filing Date: 1995-06-13  
 Number of SEQ ID NOS: 29  
 SEQ ID NO: 8  
 LENGTH: 1786  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: polypeptide  
 us-08-973-462-8

Query Match 5.0%; Score 200.5; DB 3; Length 1786;  
 Best Local Similarity 20.5%; Pred. No. 0.00018; Mismatches 329; Indels 301; Gaps 52;  
 Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;

QY 6 AES--KFKN LGNGKEGLKKDT-----TGVEHHHOENE-----ESIK 40  
 Db 868 AESVTFSNILETOENTITNDTEEKEELHENWLAALENTQSEEKEKVIDVIEVK 927





QY 316 TTYIADSRNVEBGRSTOSVUMSALDGNIITRYQVFPTFMDKGAEATDKDGINLTVDSSKLV 375

Db 203 ATI-----LNIGNIKHPSIVMT-----GLFTDDGAQV 231

QY 376 LIG-KDPKEYTGRDKNEVAKEDGSMFIFDTPKPNLMSM-DKNYENPSKSNKITYRNPB 432

Db 232 EWPPTEDISYSGCTIDNGALNBEG-----TRAKNPLPLNSGAGPATGNNSNVTKNT 285

QY 433 F--YLRKISDKGGFNWBLRNESVUDNYLIYGDLDHIDNTRDFNKLNUVKGDM--D 486

Db 286 EKDSYQCHA1QTAGSKN-----VLVNDNSRELGQA-----LOKTMKDQQLISKE 329

QY 487 WGMKYDVKANGFPDKVTDNG---NVYLQTY---SDIANK---AVGPHYOPFLDNVKE 536

Db 330 IQIBPLTRKGFYALND-DGKKSBNVTIONSYFGKSDKSGELVTAIGHYQTLSTONPSN 388

QY 537 VNDPKGNTSISYADGKSVFVNINDKENINGEGERIOCHY-----INCKEYTSEND 588

Db 389 IKLNNHHFDNNMMYAGVRFGTGFTDVLIKNRFDKVKGKESVHYRESGALVNWASYKNTD 448

QY 589 I---KOLI-----DKTUNIKUWKDAR----INTVKEILNLKDGESELK 628

Db 449 LUDLNKQVVIENIENIADPKPKAIR-VAKOABYLVKUSITVUTKVUNINNKEPO- 505

QY 629 PRHTVTUQNGKMSSTIV---SBEDFLPILPVYKGLB---EKSYQFDGWEI---SGFEGK 678

Db 506 PABELLVRSDNLVWSSENSIPGSKEGIVIEDSKGKITCHLANNQYFLNSKXVISFIKSANGK 565

QY 679 K-----DAGYYVNLSDOTPIKEVFKLEEKREE-----NKPTFDVSK 752

Db 566 EPVIRDSDGNFNT-VTENGLYKLVTNNLSDKNEKEKEKEKOYNSNAVIDSNQNGEFNS 624

QY 717 KDNPQVNHSQLINESHRK-----EDLQREEHQSKDSTKDV 752

Db 625 SKDRNQMDKDKNQDNCTEENYKIVGDGRTENHINKSKIVDV 670

RESULT 11

US-08-728-470-10 Application US/08728470

; Sequence 10, Application US/08728470

; Patent No. 5928651

## GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: Of No. 5928651-Typeable Haemophilus

NUMBER OF SEQUENCES: 10

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

## COMPUTER READABLE FORM:

MEDİUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/728,470

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

## ATTORNEY/AGENT INFORMATION:

NAME: Berkstrasser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-633

## TELECOMMUNICATION INFORMATION:

TELEFAX: (703) 415-0810

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1529 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-728-470-10

Query Match 4.7%; Score 187.5; DB 2; Length 1529; Best Local Similarity 20.1%; Pred. No. 0.001; Mismatches 255; Indels 239; Gaps 37; Matches 148; Conservative 94; MinMatches 255; Index 239; Gaps 37;

Db 511 RFPNNSVINGTGK-----GLKFIAQNQNPFTKEDGELANIS-GVTINGNTKKVW 580

QY 64 KLUKKKRREVDDFTSETGKRMEEYDY-KYDKGNITAYDGDIDLEYETEKLDEIKSYG 122

Db 581 NASKDSYWNVNSLTLNT--VOKFTPKFVDSGS-----NSQDLRSSRRSPAG 625

QY 123 VLSPSKGKHF-ILGKS-S-NVSKNARYY-----GNNYKSIET----KATYDHSKT 169

Db 626 V-----HFNGIGKGNFNGANAKLFLKLPKAATDPKBLPTTANITAGNSDS 678

QY 170 MTFDLVANIND---IVDGLLAGFDMLFLVNDOKKAELKIRMPPEKIKETKSEPYVS 224

Db 679 WMDIHANLTSRAAGINMDNSNITGDFDSITSNHNSNAEBIKOLTINATGSNS--- 735

QY 225 SYGVNVLIGEGLDLSKONPDNLTJMEEGKSYKIDUSEKOCYLLKONTILRGVALKVTTNPQ 284

Db 736 -----LKQTKSFNFYNSKHAINSSHNLTI----- 760

QY 285 KTDMLSENGVSKEDIAKIQKANPNAFALSETTYIASRNTEDGRSTOSVUMSALDGFI 344

Db 761 LGSNTVLLGGENSSSNTGNTNKTQDTSNTGNGLKTTLG----- 809

QY 345 IRYQVFPTFKMDKGAEALKDGSNLVTDSSKLVLFGK---DDKEYGEDKVNVEAKEDGS 400

Db 810 -----NISVEGULSLTGAGANANIVGNSIAESTPKGB----- 841

QY 401 MLFIDTTPKVNLSMDKNTFNPSKSNKITYVRNPBFYLGKISDKGGENWBLRNESVUDYL 460

Db 842 -----ASDNLNITGTTNGTAN-INIKQGVVKLQODDINNGGLN-ITNASGQKTI 892

QY 461 IYGLDHLNTROTPNIKANV-DGDIMWG-----MDYKAN-----G 496

Db 893 INGNT-TNEKGDLNIT-NIKADAEOIAGGNTSOKEGNLITSSDKVNTNQITIKAGVEGG 950

QY 497 FPDKVTDMDGNVLT-----GY-----SDL-----NAKAVVHY 526

Db 951 RDSSEBANANTIQTKEKLQAGDLNISGENKABTAKNGSDLTGAGNADAKVY-- 1008

QY 527 QFLYDNTRPKENVTPKNTSISYADGKSVFVNINDKRNNGFD-GETQEHYINGSKYT 584

Db 1009 --TFDKV-----DSKST-----DGHNTVNLSEVKTSNSGNSNAGDNSTGTLTSAKDT 1056

QY 585 SFDIKOIDKLNKUWKPARNTVKE-FILNKDGTGESELKPTKRVUTQONG----- 639

Db 1057 VANNVTS--HKYTNTISA---AGNVTKEGTINATGTSVE-----VIAQNGIK 1102

QY 640 -KEMSSSTIVSERDFI 653

Db 1103 NITSQNTVTTAENLV 1118

RESULT 12

US-08-719-641-10  
; Sequence 10, Application US/08719641  
; Patent No. 6218141  
; GENERAL INFORMATION:  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; NUMBER OF SEQUNCIOS: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/719,641  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-SEP-1994  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REFERENCE DOCKET NUMBER: 22,651  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1529 amino acids  
; TYPE: amino acid  
; STRANEDNESS: single  
; TOPOLOGY: linear  
; US-08-719-641-10

Query Match Best Local Similarity 4.7%; Score 187.5; DB 3; Length 1529;  
Matches 148; Conservative 94; Mismatches 255; Indels 239; Gaps 37;

QY 9 KPKNL--GNGEGSLSKKDTGVEHHQHQENRSIERSIKSSFTIDRNISTIRDPENKDLK- 63  
Db 531 RPNNSVNGTQK-----GLKFTIANQNFTHKFDGBLNIS-GIVTINOTKKDVKYW 580  
Qy 64 KLIKKEREVDFITSERGKRMEVDY-KYDKGNIAVYDDSTDLEYETEKUDBIKSYIG 122  
Db 581 MASKDSYWNVNSITLTNT--VQKPTTRKFVSGS-----NSQDIRSSRSFAG 625

QY 123 VLSPSKGHFF-ILGKTS-NVSKNAKVY----GNNYSEI---KATKVDHSKT 169  
Db 626 V-----HFNGIGKKNFNGANAKALFKPNAATDPKELPFPNANTATGNSDS 678  
Qy 170 MFDLYANIND----IVDGLAGFAGDMRLFVKNDOKKAECRKIRMPERKIKTKSEPYVS 224  
Db 679 WMDFHANLTSRAAGINMDSINTGGLDFSTISHNRNSNAPBIKKOLTINAGSNSF-- 735  
Qy 225 SYGNVIELGEJGDLSKNRPDNLTOMESKIVDSEKOYLILKRGVALKVITYNG 284  
Db 736 -----LKQTKDPSFYNEVSKHAINSSHNLT----- 760

QY 285 KYDMLLGNGVSKEDAKTOORGANPMLRALSSTTIVADSRNVEDGSTQSIVMSALDGFNI 344  
Db 761 ---LGAVNTLGGENSSSSITGNNININKANVTLQADTSNTGLKRTLTG----- 809  
Qy 345 IRYQVFTRKMDGKEDKDNLYTDSKULFGK---DKEYGDBKENVEATKEDGS 400  
Db 810 -----NTSVEGLSLTGANANIVGNLSTAEDSTFKGE----- 841  
Qy 401 MLRIDTPVNLSMDKMYFNPSPSKNSKIVYRVPPEYPLRGKISDKGGFWELRVNESVUDNYL 460  
Db 842 -----ASDNINITGTTNGTAN-INKQGWVLUQGDINNKGGIN-ITTNASGQTKI 892  
Qy 461 YGDLHIDNTRDFNIKANVKGDDIMDWG-----MDYKAN-----G 496  
Db 893 INGNI-TNEKGDLNK-NIKADAEOIGGNTSQEGNLTISSDKNNTQITIKAGVEGG 950  
Qy 497 FPDKVUTMDGVYLOT-----GY-----SDL-----NAKUVGHY 526  
Db 951 RSDSSEANANITIONKREKLAGDNTISGFKNAETAAKNGSDLTIGNASGGNADAKV-- 1008  
Qy 527 QFLYDNKPEVNIPDKPONTEYADQKSVVFNINDKRNNGFD--GRIQEHOIYINQKEYT 584  
Db 1009 --TFDKTK----DSKST----DCHNVTLNSEVTKNSNSNAGNDNSTGLTSAKDVT 1056  
Qy 585 SPNDIKQIQLDKUNIKIVVKDOPARNTVKS-FILNKOTGESELKPHRVTVIONG--- 639  
Db 1057 VNNNTVS--HKINTNSA---AGNTKEGTINATGSEV-----VTAQNGTIG 1102  
Qy 640 --KEMSTTIVSBEDFI 653  
Db 1103 NITSQNTVITATENLV 1118

RESULT 13  
US-09-200-650E-3  
; Sequence 3, Application US/09200650E  
; Patent No. 6680195  
; GENERAL INFORMATION:  
; APPLICANT: Patti, Joseph M.  
; APPLICANT: Foster, Timothy J.  
; APPLICANT: Hook, Magnus A.C.O.  
; APPLICANT: Eidhinn, Deirdre Ni  
; APPLICANT: Perkins, Samuel L.  
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from *Staphylococcus aureus*  
; FILE REFERENCE: P06283US2/BAS  
; CURRENT APPLICATION NUMBER: US/09/200,650E  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,815  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: 60/098,427  
; PRIOR FILING DATE: 1998-08-31  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 930  
; TYPE: BRT  
; ORGANISM: *staphylococcus aureus*  
; US-09-200-650E-3

Query Match Best Local Similarity 4.6%; Score 186.5; DB 4; Length 930;  
Matches 175; Conservative 121; Mismatches 346; Indels 199; Gaps 42;

Qy 3 GBAEKKFKLNLGNGREGSLKK-DTTCVHHQ--ENEESIKESSFTIDRNISTIRDDE 58  
Db 56 GELNQSKNETTAPSENKTTKVKDQHLDQPKWTMSDATVKEETSSMQSQ 115  
Qy 59 NDKLKLKCKKFREVDDFTBTGKRMEEVYDQDKGNIT-AVDGGTDLEYETEKUDIEK 117  
Db 116 NATANOSTTKSVNTINDKSSITYNE----TDKSNLTOAQDWTSPKTTIKPRTLN 169  
Qy 118 SKIYGTLSPSKDGHFBLGKISNVSKNAKYTYGNNYKSIKATKDYFHKMTWHDLYAN 177

Db 170 RMAVNTVAAPQQG-----TNV-NDKVHFSN---IDIAIDKGHVQNQTYGKTERWAT 215 ;  
 Qy 178 INDI-----VDSLAFAGDHLRFLVKNDQPKAKTAKIRMPKIKERKSKPYVSSYGVN 229 ;  
 Db 216 SSDVVLKLUKANYTIDDSVKEGD--TFTFKXGQFRPGSVRLP---SQT ONLY--NAQGNT 267 ;  
 Qy 230 ISLGEGLDLSKNEP-----DLNLTKMSESKYISDSBKKQQYLKDNLRLRGYALKVTT 281 ;  
 Db 268 IAKGIVYSTTNTTTFINNYVQYNTRG----SPEQVAFK----RNATTOKTAY 316 ;  
 Qy 282 INFQKTDMLLEGNGVYSKEDIAKIQKANPNLRASETTYAD----SRNV----EDCRST 331 ;  
 Db 317 --KMEVTLGNDTYSEETI--UDYGNKKAAQPLISSTNINNEDLSRNTAYUNQPKVYT 371 ;  
 Qy 332 QSYTLMMSALDG----NITYRVYFTEKMDKGKAIDKOGNLVMDPSSKLULRGKDKEYTGE 387 ;  
 Db 372 KOPFVNULTGYKCFPNPKNAKNFKL--YEVTDONGFVD--SFTPDTSKL----KDVNT 417 ;  
 Qy 388 DKENVEAIKEOSMLFDTKPFLNSMOKNKF-----NPSSKSNSKIVYRNPEPYLRKI 439 ;  
 Db 418 DDFDV-1YSNDNKTATVDLMIKQGOTSSNKQYIIOQVAYPDNSSTDN-----GKI 464 ;  
 Qy 440 -----SDKGKGFNWEE--LRVNTSVVDNLILYQDHLIDNTRDENIKLUVKD3-DIMWGM 489 ;  
 Db 465 DYLDTDKTIVKYSNSNSYNSSTAN--GDQKKNLGLDVWEDINKDQDANBKGI 520 ;  
 Qy 490 KOY-----KANGFP--DKYTMQGCVVLYOTGYSDINAKAVGVL-----YOFLYDNVK 534 ;  
 Db 521 KGVVVILKDSNGKELDRITTDENGKYOPTGLSN-----GTYSVEFSTPAGYPTPTANVG 574 ;  
 Qy 535 PEVNIDPKGNTS--IEVADG-----KSVVNINMK--RNNFEGDEIQRHOHYINGK 581 ;  
 Db 575 TDDAVDSGDLTTGIVKODAAMNTLDGSFYKTPKSYLDYVWTDNSKODKRSTEKGJIKGV 634 ;  
 Qy 582 EYTSFNDTQKQDIDTLNKKIVVWDKPARNTTVEPIFLNKDTGESEL--KPHRTVTION- 638 ;  
 Db 635 KVTLQNLQKGEVIGTT-----ETDENGKORFDLUDSGKVKVIFERKPAQLTGCTNT 684 ;  
 Qy 639 -----GKEMSTTIVSEEDFLIPVYKGELESKYQDFGWRSISGFECKDACYVINSKD 690 ;  
 Db 685 TPDKDADGGEVDVTTDHDF-----TLQGYVBBTSDSDSDSDSDSDSDSDSDSD 737 ;  
 Qy 691 TEIKPVVKKIEKKEERENKPKTFDVKCKDNPQVNHSQNLNESEHSRKEDIQREHNSQSKSTK 750 ;  
 Db 738 S-----DSD 788 ;  
 Qy 751 D 751 ;  
 Db 789 D 789 ;  
  
 RESULT 14 ;  
 US-09-200-650E-5 ;  
 ; Sequence 5, Application US/09200650E ;  
 ; Patent No. 6660195 ;  
 ; GENERAL INFORMATION: ;  
 ; APPLICANT: Patti, Joseph M. ;  
 ; APPLICANT: Foster, Timothy J. ;  
 ; APPLICANT: Hock, Magnus A.O. ;  
 ; APPLICANT: Eichman, Deirdre Ni ;  
 ; APPLICANT: Perkins, Samuel L. ;  
 ; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus ;  
 ; FILE REFERENCE: P06283052/BAS ;  
 ; CURRENT APPLICATION NUMBER: US/09/200,650E ;  
 ; CURRENT FILING DATE: 1998-11-25 ;  
 ; PRIORITY APPLICATION NUMBER: 60/066,815 ;  
 ; PRIORITY FILING DATE: 1997-11-26 ;  
 ; PRIORITY NUMBER: 60/098,427 ;  
 ; NUMBER OF SEQ ID NOS: 23 ;  
 ; SOFTWARE: PatentIn Ver. 2.0 ;  
 ; SEQ ID NO 5 ;  
 ; LENGTH: 1315 ;  
  
 ; TYPE: PRT ;  
 ; ORGANISM: *Staphylococcus aureus* ;  
 ; US-09-200-650E-5 ;  
 Query Match 4.6%; Score 186; DB 4; Length 1315;  
 Best Local Similarity 20.8%; Pred. No. 0-001; MilMatches 328; Indels 274; Gaps 51;  
 Matches 188; Conservative 115; MiMatches 328; Indels 274; Gaps 51;  
 Qy 23 KOTTCYEHHRHQNE-----ESTKEKSFTIRPNISTRFENKDL--KK 64 ;  
 Db 164 KSVVNVOPTNEENKKVYAKTESTTINVKSDAISNSNDETLVNN-SNSNNENNADILPKS 222 ;  
 Qy 65 LTKKPF-----REVDD-FISBTGKRMEEYDVKDDGN IIAYDDGIDL 106 ;  
 Db 223 TAPKRMTRMIAAVQPSSTEAKNVLITISNTTLTWMD----ADKNNKIVPAQDYLSL 277 ;  
 Qy 107 EYETEKUDIEK-----KIVGVLSPSKDGHFELGKISNSKNAKVVGNYK 154 ;  
 Db 278 KSQITVDKVKSGDYFTIKSYDITVQY-G-LNPE----DIKNI--GDIKOPNNGE 324 ;  
 Qy 155 SIBIKAKTYDFHSKTMF--DLYANHINDIVDGLAF-----GDMRLFWK-DN 198 ;  
 Db 325 T--IATAKHDITANNLITYTFTDYDRENNSVOMGINSYIMADTIPSKNDVEFNVTIGN 382 ;  
 Qy 199 DOKKAETKIRMPBEKIKETKSY-----PYVSYGN-----VIELRGDLISKNK- 241 ;  
 Db 383 TTUKTIANIOPVEDYVNEKNSIGAFTEVSHVGNKHNENPGYKQTIYVNPSENLSLNKL 442 ;  
 Qy 242 -----DLNLTKMSESKYSD--SEEQQYLKDNLRLRKY-----ALKV-----ALKV 278 ;  
 Db 443 KVOAHHSYSPNNI-----GQINQDVTDLKIKYQVPKYVTLNGYDWTKELTDTVNOYLK 497 ;  
 Qy 279 TTYNPGKTDMLBNGVYYSKEDIATORKANPNLRASETTYADSRNVEDGRSTQSIVMSA 338 ;  
 Db 498 ITYGDNNSAVID---FGNADSAYVVMWNTKFO-----YTN-----ESPTLVOMATLSS 543 ;  
 Qy 339 LDGFNITRYQFTKMDKGKA--IDKDGMLV-TD SK-----LVLFGK 379 ;  
 Db 544 TGNKSVSTGNALGFTNNQSGGAGQEVYKIGNYVWEDTNKONGVOELBGKGVGNUTVTFD 603 ;  
 Qy 380 DDKEYTQEDKFNVEAIKGDSMLFI----DTKPVNLIMSKDNF-KFNSKSNKIVYRNP 433 ;  
 Db 604 NTNTKQGE-----AVTFEDGSLIPNLPNGYVREBNSNLPRGYEVTEFSKQD---INNEEL 654 ;  
 Qy 434 YLRGKLSDKGGNWNELRVNESVVDNTYIYGDAHDTMRDENFTKLANYKGDIDMWGMRDYK 493 ;  
 Db 655 -----DSNGLSSVITVNGK--DN--LSADLG-----YKPKYNL--GDTV-W--EDTN 693 ;  
 Qy 494 ANGFPDK-----VTDMDGNTYLTQGYSDDANKAVWYQPL-YDVKPKEVNID 540 ;  
 Db 694 KNGIQDDBKGJISGVTVTLKOBENGNY-LKTVTTDAQK----YKFTDLODNGVYFETT 747 ;  
 Qy 541 PKGNTSTIYEADGKSVVNINNDKRN---NGDGEBOEQHIVINGKETSFENDIKQIDKT 596 ;  
 Db 748 PEGYTPITVSGSDIEKDNGLTGTINGADNMNTLDSGFYKTPK---YVGNVWMDT 803 ;  
 Qy 597 INIKIVVDFARNTTVEPIFLNKDTGEVSELKP-HRTVTIONGKEMSTIVSEEDFLP 655 ;  
 Db 804 -----NKDGKDPSTEKGISGVTVTLK-----ENGEVLO 832 ;  
 Qy 656 VTKGELLKQYDGMWESGF--EGKIDAGYVNLSTKTFEPKVKKEBKEGREENKPTD 713 ;  
 Db 833 TTKTDKQKCYOPTGLENGTGVFETPSGYT-----QVGSCTDEGIDSNGSTTG 884 ;  
 Qy 714 VSKKDKDPMQVNHSQNLNESEHSRKEDQREBHSOKSDSTKD----TATVLDKNNISSKST 767 ;  
 Db 885 VTKDKDNTIDSFGYKPTYLNQDYYWEDTNKQGVDQDKEGKISGVTVTLKODENDVLT 944 ;  
 Qy 768 TNPN 772 ;  
 Db 945 TTDEN 949 ;

Page 10

Copyright (c) 1993 - 2005 Compugen Ltd.

### OM Protein - protein search, using sw model

Run on: February 16, 2005, 19:19:57 ; Search time 141 Seconds

1791.327 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KUGBIASFKPNLNGKEGS.....ATVLDKNNTSSKSTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cggn\_6/ptcdat/a/2/pupaa/US07\_PUBCOMB.pep:\*

2: /cggn\_6/ptcdat/a/2/pupaa/PCT\_NEW\_PUB.pep:\*

3: /cggn\_6/ptcdat/a/2/pupaa/US06\_PUBCOMB.pep:\*

4: /cggn\_6/ptcdat/a/2/pupaa/US07\_NEW\_PUB.pep:\*

5: /cggn\_6/ptcdat/a/2/pupaa/US07\_NEW\_PUB.pep:\*

6: /cggn\_6/ptcdat/a/2/pupaa/PCTUS\_PUBCOMB.pep:\*

7: /cggn\_6/ptcdat/a/2/pupaa/US08\_PUBCOMB.pep:\*

8: /cggn\_6/ptcdat/a/2/pupaa/US09\_PUBCOMB.pep:\*

9: /cggn\_6/ptcdat/a/2/pupaa/US09A\_PUBCOMB.pep:\*

10: /cggn\_6/ptcdat/a/2/pupaa/US09B\_PUBCOMB.pep:\*

11: /cggn\_6/ptcdat/a/2/pupaa/US09C\_PUBCOMB.pep:\*

12: /cggn\_6/ptcdat/a/2/pupaa/US09\_NEW\_PUB.pep:\*

13: /cggn\_6/ptcdat/a/2/pupaa/US10\_PUBCOMB.pep:\*

14: /cggn\_6/ptcdat/a/2/pupaa/US10B\_PUBCOMB.pep:\*

15: /cggn\_6/ptcdat/a/2/pupaa/US10C\_PUBCOMB.pep:\*

16: /cggn\_6/ptcdat/a/2/pupaa/US10\_PUBCOMB.pep:\*

17: /cggn\_6/ptcdat/a/2/pupaa/US10\_NEW\_PUB.pep:\*

18: /cggn\_6/ptcdat/a/2/pupaa/US11\_NEW\_PUB.pep:\*

19: /cggn\_6/ptcdat/a/2/pupaa/US60\_NEW\_PUB.pep:\*

20: /cggn\_6/ptcdat/a/2/pupaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
------------	-------	--------------------	-------	-------------

; Description

RESULT 1

US-10-067-385-8

; Sequence 8, Application US/10067385

; Publication No. US200210562A1

; GENERAL INFORMATION:

; APPLICANT: Adamou, John

; APPLICANT: Choi, Gil

; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines

; FILE REFERENCE: 492901-589

; CURRENT APPLICATION NUMBER: US/10/067,385

; CURRENT FILING DATE: 2002-02-05

; PRIORITY APPLICATION NUMBER: US/09/590,991

; PRIORITY FILING DATE: 2000-06-09

; PRIORITY APPLICATION NUMBER: US/60/138,453

; SOFTWARE: PatentIn Ver. 2.0

### ALIGNMENTS

QY	1 KUGBIASFKPNLNGKEGS.....ATVLDKNNTSSKSTNNPNK 70
Db	1 KUGBIASFKPNLNGKEGS.....ATVLDKNNTSSKSTNNPNK 60
QY	61 DJKKLIKKEFREVDDEPSETGKRMEEVDYKDKGNITAYDDGTIDLEYETEKDEIKSKI 120
Db	61 DJKKLIKKEFREVDDEPSETGKRMEEVDYKDKGNITAYDDGTIDLEYETEKDEIKSKI 120
QY	121 YGVLSPKDGHFBLGKISNVSKNAKYQGYNNYKSIEIKATKDHFHKTMTFDLYANIND 180

Sequence 3, Appli  
Sequence 2, Appli  
Sequence 53254, A  
Sequence 150986,  
Sequence 2, Appli  
Sequence 52434, A  
Sequence 2070, AP  
Sequence 5719, AP  
Sequence 12751, A  
Sequence 11081, A  
Sequence 52328, A  
Sequence 52109, A  
Sequence 3792, AP  
Sequence 43, Appli  
Sequence 46664, A  
Sequence 5719, AP  
Sequence 12751, A  
Sequence 11081, A  
Sequence 76863, A  
Sequence 7309, AP  
Sequence 76866, A  
Sequence 45123, A  
Sequence 3, Appli  
Sequence 1586, AP  
Sequence 3, Appli  
Sequence 5, Appli  
Sequence 44324, A  
Sequence 254, APP  
Sequence 76871, A  
Sequence 62947, A  
Sequence 22285, A  
Sequence 20, Appli  
Sequence 53247, A  
Sequence 31, Appli

Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 YGVLSPSKDGHFELIGKISNVSKNARYTGYNNKSEIKATKDFNSKTMFDLYANIND 180  
 Qy 181 IVDGLAFAGDMLFLVKNDQKAEIKRMPBKETKSEPYVSSYGVNLGEGLSKN 240  
 Db 181 IVDGLAFAGDMLFLVKNDQKAEIKRMPBKETKSEPYVSSYGVNLGEGLSKN 240  
 Qy 241 KPDNLTKMEGSKYSDSEKOYLKONILRKGYALKVTTPKGKDMLGNGVYSKEDI 300  
 Db 241 KPDNLTKMEGSKYSDSEKOYLKONILRKGYALKVTTPKGKDMLGNGVYSKEDI 300  
 Qy 301 AKIOKANPRLASETTYADSRSNVEDGRSTSOSVMSALDFNIRIYQVFTKMDKGEA 360  
 Db 301 AKIOKANPRLASETTYADSRSNVEDGRSTSOSVMSALDFNIRIYQVFTKMDKGEA 360  
 Db 361 IDKGNLVTDSKLVLFGKDKDKEYTECDKFNATEAKDGSMLPDTKPVNLSMDKNYFNP 420  
 Qy 361 IDKGNLVTDSKLVLFGKDKDKEYTECDKFNATEAKDGSMLPDTKPVNLSMDKNYFNP 420  
 Db 361 IDKGNLVTDSKLVLFGKDKDKEYTECDKFNATEAKDGSMLPDTKPVNLSMDKNYFNP 420  
 Qy 421 SRSNKIVRNPFYLRGKISKCGPWNELRVNESVVDNYLYGDLHDNTRDPNKLNVK 480  
 Db 421 SRSNKIVRNPFYLRGKISKCGPWNELRVNESVVDNYLYGDLHDNTRDPNKLNVK 480  
 Qy 481 DGDIMDGKMDYKANGFPDKTMDGWNVYLTQGSDLNAKAVGHYQFLDNVKPEVNID 540  
 Db 481 DGDIMDGKMDYKANGFPDKTMDGWNVYLTQGSDLNAKAVGHYQFLDNVKPEVNID 540  
 Qy 541 PKGNTSLEYADGSKVWNFINDKRNGFDGETOEHITYNGKEYTSFNDIKOIDKTLNK 600  
 Db 541 PKGNTSLEYADGSKVWNFINDKRNGFDGETOEHITYNGKEYTSFNDIKOIDKTLNK 600  
 Qy 661 LRGKYODGWESGFECKKDGYVINSKOTPIKPKFKEKKELENKPKFDVSKDN 720  
 Db 661 LRGKYODGWESGFECKKDGYVINSKOTPIKPKFKEKKELENKPKFDVSKDN 720  
 Qy 721 P0VNHSLNESHRSKEDLQREBHSQKSSTDKVATVLDKONISSSTTNNPK 773  
 Db 721 P0VNHSLNESHRSKEDLQREBHSQKSSTDKVATVLDKONISSSTTNNPK 773

RESULT 2  
 US-00-769-744A-28  
 ; Sequence 28, Application US/09769744A  
 ; Publication No. US20030134407A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Le Page, Richard WP  
 ; APPLICANT: Wells, Jeremy M  
 ; APPLICANT: Hanniffy, Sean B  
 ; APPLICANT: Hansbro, Philip M  
 ; TITLE OF INVENTION: Proteins  
 ; FILE REFERENCE: PRC/P21122WO  
 ; CURRENT APPLICATION NUMBER: US/09/769-744A  
 ; CURRENT FILING DATE: 2001-01-25  
 ; PRIOR APPLICATION NUMBER: PCT/G399/02452  
 ; PRIOR FILING DATE: 1999-07-27  
 ; PRIOR APPLICATION NUMBER: GB 9816336 .3  
 ; PRIOR FILING DATE: 1998-07-27  
 ; PRIOR APPLICATION NUMBER: US 60/125329  
 ; NUMBER OF SEQ ID NOS: 196  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO: 28  
 ; LENGTH: 2119  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus pneumoniae  
 ; US-09-769-744A-28

Query Match Local Similarity 100.0%; Score 4026; DB 10; Length 2119;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-216; Length 2119;

---

Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1313 KLGELAESFKNGLNGKREGSLSKCDTGVERHHQENBESIEKKSSTIDRNISTRD芬K 60  
 Qy 61 DKKLKKKREVDPTSEGKMRMERYDYKDKNITIATDGTLEYETEKLDEIKSI 120  
 Db 1373 DKKLKKKREVDPTSEGKMRMERYDYKDKNITIATDGTLEYETEKLDEIKSI 1432  
 Qy 121 YGVLSPSKDGHFELIGKISNVSKNARYTGYNNKSEIKATKDFHSKMTFDLYANIND 180  
 Db 1433 YGVLSPSKDGHFELIGKISNVSKNARYTGYNNKSEIKATKDFHSKMTFDLYANIND 1492  
 Qy 181 IVDGLAFAGDMLFLVKNDQKAEIKRMPBKETKSEPYVSSYGVNLGEGLSKN 240  
 Db 1493 IVDGLAFAGDMLFLVKNDQKAEIKRMPBKETKSEPYVSSYGVNLGEGLSKN 1552  
 Qy 241 KPDNLTKMEGSKYSDSEKOYLKONILRKGYALKVTTPKGKDMLGNGVYSKEDI 300  
 Db 1553 KPDNLTKMEGSKYSDSEKOYLKONILRKGYALKVTTPKGKDMLGNGVYSKEDI 1612  
 Qy 301 AKIOKANPRLASETTYADSRSNVEDGRSTSOSVMSALDFNIRIYQVFTKMDKGEA 360  
 Db 1613 AKIOKANPRLASETTYADSRSNVEDGRSTSOSVMSALDFNIRIYQVFTKMDKGEA 1672  
 Qy 361 IDKGNLVTDSKLVLFGKDKDKEYTECDKFNATEAKDGSMLPDTKPVNLSMDKNYFNP 420  
 Db 1673 IDKGNLVTDSKLVLFGKDKDKEYTECDKFNATEAKDGSMLPDTKPVNLSMDKNYFNP 1732  
 Qy 421 SRSNKIVRNPFYLRGKISKCGPWNELRVNESVVDNYLYGDLHDNTRDPNKLNVK 480  
 Db 1733 SRSNKIVRNPFYLRGKISKCGPWNELRVNESVVDNYLYGDLHDNTRDPNKLNVK 1792  
 Qy 481 DGDIMDGKMDYKANGFPDKTMDGWNVYLTQGSDLNAKAVGHYQFLDNVKPEVNID 540  
 Db 1793 DGDIMDGKMDYKANGFPDKTMDGWNVYLTQGSDLNAKAVGHYQFLDNVKPEVNID 1852  
 Qy 541 PKGNTSLEYADGSKVWNFINDKRNGFDGETOEHITYNGKEYTSFNDIKOIDKTLNK 600  
 Db 1853 PKGNTSLEYADGSKVWNFINDKRNGFDGETOEHITYNGKEYTSFNDIKOIDKTLNK 1912  
 Qy 601 IVDGLAFAGDMLFLVKNDQKAEIKRMPBKETKSEPYVSSYGVNLGEGLSKN 240  
 Db 1913 IVDGLAFAGDMLFLVKNDQKAEIKRMPBKETKSEPYVSSYGVNLGEGLSKN 1972  
 Qy 661 LRGKYODGWESGFECKKDGYVINSKOTPIKPKFKEKKELENKPKFDVSKDN 720  
 Db 1973 LRGKYODGWESGFECKKDGYVINSKOTPIKPKFKEKKELENKPKFDVSKDN 2032  
 Qy 721 P0VNHSLNESHRSKEDLQREBHSQKSSTDKVATVLDKONISSSTTNNPK 773  
 Db 2033 P0VNHSLNESHRSKEDLQREBHSQKSSTDKVATVLDKONISSSTTNNPK 2085

RESULT 3  
 US-10-282-122A-73670  
 ; Sequence 73670, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

PFILE\_REFERENCE: ELTRA\_034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; PRIORITY APPLICATION NUMBER: 60/111,078  
; PRIORITY FILING DATE: 2000-03-21  
; PRIORITY APPLICATION NUMBER: 60/206,848  
; PRIORITY FILING DATE: 2000-05-23  
; PRIORITY APPLICATION NUMBER: 60/207,727  
; PRIORITY FILING DATE: 2000-05-26  
; PRIORITY APPLICATION NUMBER: 60/230,335  
; PRIORITY FILING DATE: 2000-09-06  
; PRIORITY APPLICATION NUMBER: 60/230,347  
; PRIORITY FILING DATE: 2000-09-09  
; PRIORITY APPLICATION NUMBER: 60/242,578  
; PRIORITY FILING DATE: 2000-10-23  
; PRIORITY APPLICATION NUMBER: 60/253,625  
; PRIORITY FILING DATE: 2000-11-27  
; PRIORITY APPLICATION NUMBER: 60/257,931  
; PRIORITY FILING DATE: 2000-12-22  
; PRIORITY APPLICATION NUMBER: 60/267,636  
; PRIORITY FILING DATE: 2001-02-09  
; PRIORITY APPLICATION NUMBER: 60/269,308  
; PRIORITY FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 73670  
; LENGTH: 2140  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
; US-10-282-122A-73670

Query Match 100.0%; Score 4026; DB 15; Length 2140;  
; Best Local Similarity 100.0%; Pred. No. 5.9e-216;  
; Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIGEIAASKFKNGLNGKGSLSKDKTGVHHEQNEESIKESSEFTDRNSTIOPENK 60  
Db 1334 KIGEIAESKFKNLNGKGEGLSKDKTGVHHEQNEESIKESKFTDRNSTIOPENK 1393  
QY 61 DJKLUKKRFREDDFSETGKMEEDYKDKGNTIAYDGTDLYEETKLDEIKSI 120  
Db 1394 DJKLUKKRFREDDFSETGKMEEDYKDKGNTIAYDGTDLYEETKLDEIKSI 1453  
QY 121 YGVLSPSKDGHFELGKISNSVSKNAVKYYGNNYKSTETKATKDHSKTMFDLYANIND 180  
Db 1454 YGVLSPSKDGHFELGKISNSVSKNAVKYYGNNYKIEKATKDHSKTMFDLYANIND 1513  
QY 181 YVDGLAGPAGDMFLVKNDQKAEIKRMPPEKIKETSEPPVSSGVNLIEGEGDSK 240  
Db 1514 YVDGLAGPAGDMFLVKNDQKAEIKRMPPEKIKETSEPPVSSGVNLIEGEGDSK 1573  
QY 241 KDPDLTKMESGKYSKDSERKQOQLKDNILRGYALKUTTNGKTDMLLEGNGVSKEDI 300  
Db 1574 KDPDLTKMESGKYSKDSERKQOQLKDNILRGYALKUTTNGKTDMLLEGNGVSKEDI 1633  
QY 301 AKIQKANPNRLASETTYIADSRAVEDGRSTOSVMSALDGENIIRYQVFTRNDKGEA 360  
Db 1634 AKIQKANPNRLASETTYIADSRAVEDGRSTOSVMSALDGENIIRYQVFTRNDKGEA 1693  
QY 361 IKDGNLVTDSKLVLKRDKEYTGEKDKNVAKEDGSMFLIDTPVNLSMDKNYFNP 420  
Db 1694 IKDGNLVTDSKLVLKRDKEYTGEKDKNVAKEDGSMFLIDTPVNLSMDKNYFNP 1753  
QY 421 SIEKNIKVNRPEYLGRKISDKGFPNWLVRVNESWVNDYLIVGLDHIDNTROPNIKLNK 480  
Db 1754 SKSNKTVYRNPEFYLRGKISDKGFPNWLVRVNESWVNDYLIVGLDHIDNTROPNIKLNK 1813  
QY 481 DGDIMDGKWDYKANGFPDKVTMDGNYLQTCYSDJNAAKAVHVKQFLYDNKPEWID 540  
Db 1814 DGDIMDGKWDYKANGFPDKVTMDGNYLQTCYSDJNAAKAVHVKQFLYDNKPEWID 1873  
QY 541 PKGNTSIEYADGKSVVFNINDKRNGFDGETOBOHIYINGKEYTSNDIKQIDKTLNIK 600

RESULT 4  
; Sequence 1180, Application US/10472928  
; Publication No. US20050020813A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SPA  
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLBIC ACIDS  
; FILE REFERENCE: P26926WO  
; CURRENT APPLICATION NUMBER: US/10/472, 928  
; PRIORITY FILING DATE: 2003-09-26  
; PRIORITY APPLICATION NUMBER: GB-0107658.7  
; PRIORITY FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 4979  
; SOFTWARE: SeqWin99, version 1.03  
; SEQ ID NO: 1180  
; LENGTH: 2140  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
; FEATURES:  
; OTHER INFORMATION: serine protease, subtilase family  
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)  
; OTHER INFORMATION: Similar to strain R6 Sequence 15902605 (0.B+01)  
; US-10-472-28-1180

Query Match 100.0%; Score 4026; DB 17; Length 2140;  
; Best Local Similarity 100.0%; Pred. No. 5.9e-216;  
; Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIGEIAASKFKNGLNGKGSLSKDKTGVHHEQNEESIKESSEFTDRNSTIOPENK 60  
Db 1334 KIGEIAESKFKNLNGKGEGLSKDKTGVHHEQNEESIKESKFTDRNSTIOPENK 1393  
QY 61 DJKLUKKRFREDDFSETGKMEEDYKDKGNTIAYDGTDLYEETKLDEIKSI 120  
Db 1394 DJKLUKKRFREDDFSETGKMEEDYKDKGNTIAYDGTDLYEETKLDEIKSI 1453  
QY 121 YGVLSPSKDGHFELGKISNSVSKNAVKYYGNNYKSTETKATKDHSKTMFDLYANIND 180  
Db 1454 YGVLSPSKDGHFELGKISNSVSKNAVKYYGNNYKIEKATKDHSKTMFDLYANIND 1513  
QY 181 YVDGLAGPAGDMFLVKNDQKAEIKRMPPEKIKETSEPPVSSGVNLIEGEGDSK 240  
Db 1514 YVDGLAGPAGDMFLVKNDQKAEIKRMPPEKIKETSEPPVSSGVNLIEGEGDSK 1573  
QY 241 KDPDLTKMESGKYSKDSERKQOQLKDNILRGYALKUTTNGKTDMLLEGNGVSKEDI 300  
Db 1574 KDPDLTKMESGKYSKDSERKQOQLKDNILRGYALKUTTNGKTDMLLEGNGVSKEDI 1633  
QY 301 AKIQKANPNRLASETTYIADSRAVEDGRSTOSVMSALDGENIIRYQVFTRNDKGEA 360  
Db 1634 AKIQKANPNRLASETTYIADSRAVEDGRSTOSVMSALDGENIIRYQVFTRNDKGEA 1693  
QY 361 IKDGNLVTDSKLVLKRDKEYTGEKDKNVAKEDGSMFLIDTPVNLSMDKNYFNP 420  
Db 1694 IKDGNLVTDSKLVLKRDKEYTGEKDKNVAKEDGSMFLIDTPVNLSMDKNYFNP 1753

QY 421 SSKSNKTYVNPYRPLRGTSKGGFNLWLRVNESWDNYLYGDLHDNTDPNKLVK 480  
Db 1754 SKSNKTYVNPYRPLRGTSKGGFNLWLRVNESWDNYLYGDLHDNTDPNKLVK 1813  
QY 481 DGDIDMPGMKDIXKANGEPDKVTDGMGAVYLQGYSILANAKVGHYQFLYNVKPENID 540  
Db 1814 DGDIDMPGMKDIXKANGEPDKVTDGMGAVYLQGYSILANAKVGHYQFLYNVKPENID 1873  
QY 541 PKGNTSIEYADGSKVWNINDRNGDGEQEHYINGKEYTSNDIKIQLDNLK 600  
Db 1874 PKGNTSIEYADGSKVWNINDRNGDGEQEHYINGKEYTSNDIKIQLDNLK 1933  
QY 601 IWKDFARNTTKEFLANKDTEVSEALKPHRTVTONGKMSSTVSEERFILPKGE 660  
Db 1934 IWKDFARNTTKEFLANKDTEVSEALKPHRTVTONGKMSSTVSEERFILPKGE 1993  
QY 661 LKGYQFDGWESISGFEGKDKAGYVILNSKDTIKPVFKKIRKEBNKPTFDVSCKDN 720  
Db 1994 LKGYQFDGWESISGFEGKDKAGYVILNSKDTIKPVFKKIRKEBNKPTFDVSCKDN 2053  
QY 721 PQNHSOLNESRKEQDQEBSQSDSTKOTATVLDKNMISKTTNNPK 773  
Db 2054 PQNHSOLNESRKEQDQEBSQSDSTKOTATVLDKNMISKTTNNPK 2106

RESULT 5  
US-03-765-272-68  
; Sequence 68, Application US/09765272  
; Patent No. US2002006155A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20855  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MS DOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/765,272  
; FILING DATE: 22-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 35,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 68:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acid  
; TYPE: amino acid  
; STRANDBENDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
; US-03-765-272-68

Query Match 5 6%; Score 227; DB 15; Length 1196;  
Best Local Similarity 20.2%; Pred. No. 0.0029%; Matches 185; Conservative 145; Mismatches 260; Indels 328; Gaps 46;

QY 1 KLGTAESKFNLGNQEG-----SIKKITGVHHHQNEEST 39  
Db 123 RUKDQF-LFENDTGKKEGSISIIGGKIAVLGKPERRSILBEEAGTVKFKWKEED 181  
QY 40 KEKSSFTIDRNISTRDFEN--KDKLKLKKFREVDFF-TSETGKRM----- 85  
Db 182 KKLSN-TEQNLIRKDLINTVEERMEPLKEESEKAKKFLNLSLEAAGTVKFKWKEED 239  
QY 86 -EVDYKDDKGNTIAYDGPD-LEYETEKDEIKSKYGLPSRSDGHFELGKSVNSVK 143  
; ORGANISM: Clostridium botulinum  
; US-10-282-122A-52737





Db 337 DKNDPIDIITDSGEVIAKTFDKATKQTYTFTDVKYEDIKARLJLYS-----IDK 390 ; PRIOR APPLICATION NUMBER: 60/191,078  
 Qy 175 YANINDIVGJAF-----GD--MRLFKV DNDQGKAIEKIRM- 209 ; PRIOR FILING DATE: 2000-03-21  
 Db 391 QAVPNETSLNLTIFATACKETSONVSVDYQDPMVHGSNTIQIIFTKUDENKOTTEODIYVN 450 ; PRIOR APPLICATION NUMBER: 60/206,848  
 Qy 210 PEKIKETKEY---PVSSYONVIENGEGD--LSKQKPDNLTKM-----ESGKYSD 256 ; PRIOR APPLICATION NUMBER: 60/205,347  
 Db 451 PLKKTATNTKDIAGSQVDDQYGN-IKLNGNSIIDONTEIKWVQNPQOIPQSNIYDF 509 ; PRIOR FILING DATE: 2000-05-23  
 Qy 257 SEKQOYILK-DMLRKGYALKVTTNGKTMLEGINGVYKEDIARIQOKANPNUALSE 315 ; PRIOR APPLICATION NUMBER: 60/230,335  
 Db 510 SOYEDVTSQFDN---KGSFSNNVATLDFG----DINSAYI-----IKVSK 548 ; PRIOR FILING DATE: 2000-09-06  
 Qy 316 TTYAIVSR-NVEDGRSTQSVMSALDOFNIRYQVFPMKMG----- 358 ; PRIOR APPLICATION NUMBER: 60/230,347  
 Db 549 YKPTSDCEBLDAQTSKRT--DKYQYNYAGYSNFTVTSIDGGDGTVPKBPEKLYIG 606 ; PRIOR APPLICATION NUMBER: 60/242,578  
 Qy 359 ---EATDKDQSNLVTDSSKLVJLFGKODKEYGEDKFNVAEIKEDGSMFLFIDTKPVNLSMD 414 ; PRIOR FILING DATE: 2000-11-27  
 Db 607 DVWEDDVKDQG-----VOGTSKESKPHANVL-VTLYTPDT-----TKSVRTDN 650 ; PRIOR FILING DATE: 2000-09-09  
 Qy 415 KNY--ENPSKSNIKLYV--RNPPRYLRKIS-----DKGFNWLVRVNESENVDNTLYG 463 ; PRIOR APPLICATION NUMBER: 60/267,636  
 Db 651 GHIEFGGLKDGTIVYKPETPGYLPTKVNGETDGEKDSNNSSITVKNG-----D 702 ; PRIOR FILING DATE: 2001-02-09  
 Qy 464 DHID-----NTRDPNITKLNVKDQG--DIMWGMKDYANGFPDKVUTMDGNTLQT 512 ; PRIOR APPLICATION NUMBER: 60/253,625  
 Db 703 DMSLDTCPYKPERKYNQGDYVWDTNKOGIDQANEPGKDVYK-----LKOSTGKV-IGT 756 ; PRIOR FILING DATE: 2001-02-16  
 Qy 513 GYSDLNAKAVGVHYQFL-YDNVYKEPVNID-PKG-----NTSIEYAD--GKSUVFNINDK 562 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 Db 757 TTIDASCK---YKFPLDNGNTVEFETAPGTYPTPKNTAEDKDSNGLJITGVIKA 811 ; NUMBER OF SEQ ID NOS: 78614  
 Qy 563 RNRNGFDEBIQBOHIYI-----NGKEYTSEFDNIKQDIDKTANIK-TVKDFAART 610 ; SOFTWARE: Patentin version 3.1  
 Db 812 DNMTLDGSPYKTPKYSLGDVWYDSNKDQKQDSTEKGKJDVKVTLNEKGIVIGITKDE 871 ; SEQ ID NO: 76365  
 Qy 611 TYKEFTLANKDGEVSEL--KPRHVTUQN-----GKEMSSITVSEBDPILPVKG 659 ; LENGTH: 4688  
 Db 872 NCKYRFDNLDSSKVKFPEKPGILOTOVTNTTDDKODADGEGDVTTDHDPI----- 925 ; TYPE: PRT  
 Qy 660 ELEKGYQPDGWEISGFECKKDAQGVINLISKOTPIKPVKKLBKKEBKENKTPDVSKKKD 719 ; ORGANISM: Ureaplasma urealyticum  
 Db 926 -LDNGYFEEDTSDSDSDSDSDS-----DSDSDSDSDSDS 975 ; US-10-282-1122A-76865  
 Qy 720 NPOVNHSQLNESRHKEDLQREEHQSQSDSTD 751 ;  
 Db 976 SPSDSDSDSDSDSDSDSDSDS 1007 ;  
 RESULT 10  
 US-10-282-1122A-76865  
 ; Sequence 76865, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Lianggu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Charyl  
 ; APPLICANT: Ohlson, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELTRA\_034A  
 ; CURRENT APPLICATION NUMBER: US/10/282.1122A  
 ; CURRENT FILING DATE: 2003-02-20

Query Match 5.2%; Score 207.5; DB 15; Length 4688;  
 Best Local Similarity 19.6%; Pred. No. 0.02; Mi. Matches 342; Indels 252; Gaps 47;  
 Matches 179; Conservative 142; Mismatches 342; DB 15; Length 4688;  
 Qy 21 LKDTDTGVEHHHQENSESSEKKSFTIDRNSTIRDEFENDKDKLKKFREVDPTSET 80 ;  
 Db 78 INKDVYCV--FVDENNQBYKQ--IDSYKAF-BFTKYLPKTCKYKLMRITDANNI 131 ;  
 Qy 81 GKRMEYDYKDDKGII-----AYDDGIDLEYETEKLDI-----KSKY 121 ;  
 Db 132 VRENEK--LKQBQKTIULKPVPTNVYEDQNEKHOLVKVNKLANKLIVAKFVNQNNEVY 189 ;  
 Qy 122 GYLSPSPDGHFEBILGKII--SNVSKRKAQVYVGNNYKSEIKAJKDFHSKMTFDIYANI 178 ;  
 Db 190 SI-----BAHVQDOKIDITSSLKQDQVY--LKDLIK-----DSNPLT--KVNN 233 ;  
 Qy 179 NDIVGJAFAGDMRLTVKDQKQKARKKIRMPKIK-----ETSEYY--PVSS 225 ;  
 Db 234 NDIDE--OPKKIDKTMVYDNGNLV-LKVIDTNTKQDQALPTAVFKDENNEYLIPALSV 290 ;  
 Qy 226 YGNVIELGEGDLSKNPKDNLTKMESSKLYSEKQOYLKONILRKGYALKVTTNGK 285 ;  
 Db 291 KNNV-----GYENTKLLPLNPKYELMI-VWNNELANDVLPNELMLE-HKVSUNKTAK 344 ;  
 Qy 286 TDMLEGNGVYSEKDIARIQKAMP--NRLAUSSETTYADSRSRVEDCGRSTSUVMSALDF 342 ;  
 Db 345 VNFDNPKVY-EVDLGVQLKNTPLETLEDLHQYKINKAKTDEKORAVFDI--SSIGDN 401 ;  
 Qy 343 NTRYOFTEKNDKGKHAID-----KOGNL-----VTDSSKLUFLGKDKDE 383 ;  
 Db 402 NL--YETGIKTKNEVUNVHQIYPHNRSRINSTLNNSALNTPYQYKNGDNLIAKAPY 459 ;  
 Qy 384 YTGBDKENV-----BAIKEGSMFLFIDTKPVNLSMDKNYFFNPSKSNKIVRNP 431 ;  
 Db 460 YVNOQVGYFKDNONBHQOLAKVGKGTAF-DRCALN----NNSNSLSDKTVVSNSP 513 ;  
 Qy 432 E-----FYLRK--1SDKGFPNWELL-----RVAESVVDYLI--YGDJHIDNT 470 ;  
 Db 514 QWVLVSNFDLISKOKOLIKKGAANASVSDSTKQTQILENLNLNDLNGLKVATPD--NND 570 ;  
 Qy 471 RDFTNIKANVKQD--DIMWGMKDYANGFPD-----KVMDGMVYLOGYSDIANKA 521 ;  
 Db 571 KEYKVVANVDONKIEFD-----SNDLPGKYTHLAKCENNDLNKVKINLDFELDKI 623 ;

Query Match 5.1%; Score 207; DB 15; Length 1191;  
 Best Local Similarity 19.6%; Pred. No. 0\_0038;  
 Matches 178; Conservative 139; Mismatches 275; Indels 314; Gaps 43;  
 SEQ ID NO 52048

QY 624 IDKRDNLNLDS-HPDFTYNDGNLTH---TOIANDLNDLKKORALNNANVKGIVWDQ- 575  
 QY 576 IYINGKETISFNDIKQIDKTLNIKIVVKDFARTIVKS---FLINKDTGEVSELKPHRV 633  
 QY 678 ---DGIEH---EIDVSIDANGKVILPTKNLANNUPTKNITYTRK----- 716  
 QY 634 VTIQNGKENSSTIVSEE-----DFILPVYKGSELEKGYQFDGMEIS----- 673  
 QY 717 WVLKQNNOPNDLISSEEQLSGDNHISFKPKEITAKTKEN--DDYEISFSNPLANKK 773  
 QY 674 -----GPEGKDAGTVINLSDTFPKPVEKKEENRPTFDV 714  
 QY 774 LTPKTDNNNTNTKTEASIGLDGKA---IFKTSDDATAPDHKT-LTKIEDANRKVANI 829  
 QY 715 SKKKDNPQNHQSOLNEBHSRKEDLOBEHHSOKSDSTKVATVLDKN----- 761  
 Db 830 DEISPLDRIVNKQNGNWNAD--KHEKFIPDQKNKOLTAVKOKNNELHVPIKTDKG 887  
 QY 762 --ISKSSTTNNPNK 773  
 Db 888 KVIVNPNNNLFDPNK 902

RESULT 11

US-10-282-122A-52048

Sequence 52048, Application US/10282122A  
 Publication No. US20040029129A1  
 GENERAL INFORMATION:

APPLICANT: Wang, Liangsu  
 APPLICANT: Zainduo, Carlos  
 APPLICANT: Malone, Cheryl  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kari  
 APPLICANT: Zyskind, Judith  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John  
 APPLICANT: Carr, Grant  
 APPLICANT: Yamamoto, Robert  
 APPLICANT: Forbyth, R.  
 APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA\_034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 52048

LENGTH: 1191

TYPE: PRT

ORGANISM: Clostridium acetobutylicum

RESULT 12

US-10-744-616-7

QY 1 KLGEBIERSKFRNGLNGKEG-----SLKKDTGVERHHHQENSEI 39  
 QY 120 RUKDGT-LFDUTGIGKCGYSTIGQGKBEAVLSGKPPERRALLAEAGIVKPKTRVDA- 177  
 QY 40 KEKSSFTIDRNISTIRD-----FENKDLKLKKFREVD--FTSETGK--RMEE 86  
 QY 178 -EKKLENQNOLVRINDLRLTYERLEPLRESEKARFVELSDELKTKKEINTIYSDN 236  
 QY 87 YDVKYDKRGNTIA---YDDCT-----DLYEYETKDOI-----KSKIVGVLSP 126  
 QY 237 IDYRINDLKQWADLKSISDENVKDEKEKISBLKATESLIDEFDAYKSSNKRY---E 292  
 QY 127 SKDGHBTGK-----SNVSKVAKVYIGNNSIE-----IKATYDFHFKMTDLY 175  
 Db 293 SKSEROKILSIELLEKKTSHSDVAKNLYKEBEDLNDSVN-KSYEIQKLTED-- 349  
 QY 176 ANINDIVGLAPAGDNRLFVQNDQKABKIRMP--KIKETSE---YPVSSVGN 228  
 Db 350 KNYNK-----ELSKINWKSEEKGKNGNLIESNEKSIQKOKDADLISTISONN 400  
 QY 229 VIELGBGDLISKKNPKDNLTKMSS---GKIYSDSEKOQYLUDNITLRKGYALKVTNPG 284  
 Db 401 EIVI---LKRKIESNESKLSKISKRAGEGYSK-----LKINEVNTKTLSELVKIN- 448  
 QY 285 KUDMULGNGVVSKEDEJAKIQKANPNURALSETTIYADSNNFEDGRSQTQVSUMSALDFN 344  
 Db 449 ---DKTSGYENGIOIRENRNISKLN-----RIISDE----- 475  
 QY 345 TRYQVFTKMDKGEALDKDQNLVTDSSKLVLFKGDKDKEYGEDKFNVAEIKDGSMLFI 404  
 Db 476 --EKLNLRLMSKSNKLEANQMLNL-----EKQYGGYNR-SVKNLMOHVTGFV 522  
 QY 405 DTKPVNULSMMDKNYFNPSKSNTKLYVRNRPFYLRGKISDKGGFWELRVNESVVD--NYLI 461  
 Db 523 DVKPESFFVLSQEVIVKVKKEFETAV--EISIGAISD-----ITLDDNIAKKLINYLK 573  
 QY 462 YGDHLHDNTRPFI---KUNVKDG----- 482  
 Db 574 SKNLGRATPLHLNIKGRKLNISDACTRHEKORIGIASELIBYDSTFLPAVNTVVLGRTV 633  
 QY 483 DJMDW&MKDYKANGFPDKVTMDGNVYLQTY3YSDLMNAKAV3VHYQFLYDNKVEPVNIDPK 542  
 Db 634 DMDSAKIKALANSYSPKIVLUTGEV----- 663  
 QY 543 GNTS--IEYADGKSVVNFNINDKRNNGPFDGETQBOHTYINGKEYTSENDIKOIDKTLNIK 600  
 Db 664 GSITGGSTSYKASTI-----GRKREBELNBLNNVSOA---LEOSSNKLKEN 710  
 QY 601 WVKDOPARNTTKEFLNKDTGEVSELKPHRVVTONGKENSSTIVSEEDTLPVYKG 660  
 Db 711 KVVKE-----LDNLCUD-----LTDI-HGEKELTKKERLKSIDISEK 750  
 QY 661 LAKGYOPDGWTSFGFEGKDAGYVINLSDKTFPKPFKREK--KEEENKPTFDVSK 717  
 Db 751 LANKSYTIVSGIG-----PIKEKINKHLKLUKXKEENKA--LKLR 789  
 QY 718 KNPQVNHQSNEHRSK-----DLQREBHSQSD---STKVATVLDK 759  
 Db 790 RAN---NNNLDELERLKKDENSKVNLNEEIMSMKVDAKSDEMMSSTRETERYKVEM 846  
 QY 760 NMISK 765  
 Db 847 HNMENK 852

; Sequence 7, Application US/10744616  
; Publication No. US20050026170A1  
; GENERAL INFORMATION:  
; APPLICANT: Parti, Joseph M.  
; APPLICANT: Foster, Timothy J.  
; APPLICANT: Hook, Magnus A.O.  
; APPLICANT: Eidhian, Deirdre N.  
; APPLICANT: Perkins, Samuel L.  
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from *Staphylococcus aureus*  
; FILE REFERENCE: P06281US2/BAS  
; CURRENT APPLICATION NUMBER: US/10/744,616  
; CURRENT FILING DATE: 2003-12-24  
; PRIORITY APPLICATION NUMBER: 60/066,815  
; PRIORITY FILING DATE: 1997-11-26  
; PRIORITY APPLICATION NUMBER: 60/098,427  
; PRIORITY FILING DATE: 1999-08-31  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 1166  
; TYPE: PRT  
; ORGANISM: *Staphylococcus aureus*  
; US-10-744-616-7

Query Match 5.0%; Score 203; DB 17; Length 1166;  
Best Local Similarity 21.6%; Pred. No. 0.062; Gaps 53;  
Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;

Qy 33 QBNBESIKE-KSSFTIDRNISTIRDRENKDLKLKKR-----EVDDFTS 78  
Db 226 KNPKEKUKELVANDSNTHDSTREVATAPTSVAKRWNKRAVQAFAAVASNNDLIK 285  
Qy 79 ETGKRMEDYKVKDKGNIIAYDGTLEYET-----KLEDEIKSKIVGVL 125  
Db 286 VTKQTIK---VGDKGONVAHAHDGKDIBYDIEFTIDNKVKKGDWTINYDKNVIFSDLT 341

Qy 126 PSKED-----GHFETLIGKISNVSKNAK---VYVGNNYKSTEIKATKYPDFHSK---- 168  
Db 342 DKNDPDPIDTDPSEGVIAKGTDFKDATKQIYTTFDYVQYEDIKSRMLYSYDKKTPWNE 401

Qy 169 ---TMTP-----DLYANIN-DIVGLAPAGD--MRLFKV-DNDQKGAEIKRM-PPEK 214  
Db 402 TSLNLTATAGETQSNTVQDPMW-GDSNIQSTFTKQJEDQKTEQOQIVNPIKKS 460

Qy 215 ETKSEY---PVYSSYGNVILEGEGD--LSKQKPDNLTRM-----ESGKYSDSBKQ 261  
Db 461 ATWTKVDIAGSQVDDYGN-IKUGNSTIDQVTEIKYKVMSDQLRQPSNRKIDFSQED 519

Qy 262 VILK-DNLTKRGYALKVTTVNGKTMLEMGVSYKEDAIKLOKAMPNRLASETTIYA 320  
Db 520 VTSOFDN---KKSFSNNVATLDFG---DINSAYI-----IKVWSKVPPTS 558

Qy 321 DSR-NVEDGRSTOSVULMSALDGENIIRYQVFPTKMDKG-----E 359  
Db 559 DEBLDIQGTSRHTT--DKYGYNNYAGCNSNFTWTSNTGGGDTVKREKLUKIGDVME 616

Qy 360 A1DKDGNLVTDSKLVFGKDKKEYTSEDKFENVEAKEDGSMFLIDTRPVNTSMDRK-Y 418  
Db 617 DVDKDG-----VOGTSKEKPMANL\_VLTIPDGT---TKSVRITDANGHIEF 660

Qy 419 NPSKSNKLYV---RNPFEPYLRSKIS-----DKGGENWELLRVNESTVNDNLYGDAHD 468  
Db 661 GGLKDGTYTVEKTFPTGYLPTKNGTIDGEKOSNGSSVTVKINGK-----DDBMSLD 712

Qy 469 -----NTDPFNKLANKVG---DIMDMGMKDYKANGFPKUTDMGNNVLTQGVSDL 517  
Db 713 TGPYKEPKYLNQDYVWEDTNKOGIQDANEPIGKIDVKY---LKOSTGKV-IGTTTDA 766

Qy 518 NAKAVGWHYQFLJDNKPEVNIDPKGNTSIEV---ADGKSVVFN---INDKRNNGFDGEIQ 572  
Db 767 SGK-----YKF-----TLDG-NGMYTVEPFTPACTPTVKTADDKOSNL--- 807

Qy 573 EOHIVYINGKEYTISFNDIKOIQIDKTLNIRKIVVKDFARNT--TVKEFI---LNKDTGEVSEL 627

RESULT 13  
Sequence 7646, Application US/10032585  
Publication No. US20030180953A1  
; GENERAL INFORMATION:  
; APPLICANT: Terry, Roemer D.  
; APPLICANT: Bo, Jiang  
; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-999  
; CURRENT APPLICATION NUMBER: US/10/7646  
; CURRENT FILING DATE: 2001-12-20  
; PRIORITY APPLICATION NUMBER: US/10/032,585  
; PRIORITY FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7646  
; LENGTH: 1881  
; TYPE: PRT  
; ORGANISM: *Candida albicans*  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (1881)..(1881)  
; OTHER INFORMATION: X=any amino acid  
; US-10-032-585-7646

Query Match 5.0%; Score 203; DB 14; Length 1881;  
Best Local Similarity 20.8%; Pred. No. 0.011; Gaps 39;  
Matches 172; Conservative 132; Mismatches 300; Indels 222; Gaps 39;

Qy 22 KDTTGYE---HHHQNEESIEKESKFTURN-1STIRDRENKDKLKKTFREWD 76  
Db 1140 KTDATDEETKTKHITDQBERAKSQQSOPESERINDKNSDNEAKELSD-NBEKLSNNEKE 1198

Qy 77 TSETGKMEEDYKIDKGNIIAY--DDGTLEYETEAKDLFISKLYGVLSPSKDGHEI 134  
Db 1199 KTELANNLKTQDEKISLRTSWAISDKSKSLXHDDELRKRIKETTLKENETMF 1258

Qy 135 LGKISNTSKAIVYGNYKSEIEKATKYPDFHSKTMFDLYANINDIVGLAFGDMRL 194  
Db 1259 KEOLQVNDCK-----BLEAC-----LKKJETKEKEINDLIRKLEA----- 1297

Qy 195 VKNDPDKKAETKIRMPPEKIKETSEPYVSSYGNVILEGEGDLSKONKPDNLTKMESGK 254  
Db 1298 KSDHDTERRKSL-----LIEDTKSE-----SBKVNKLNE-----QBEKL 1334

Qy 255 SDSEKOQULLKONTILRKGYALKVTTVNGKTM-----LEGNGVSYKEDAIKIQ 305  
Db 1335 GEREKE---VRD---IOSLAQAKTTWWEKIITDVKLKEKSDLEKT---NKESVDTLKK 1385

Qy 306 ANPNLR---ALSETTIVADSRNVEDGSTQSULMSALDGFNI-----IRQVQF 350  
Db 1386 EVENLKEIISLLEDQKDDTICKYKELAQLETKTSNDSTIMELEKTELEKKVNLTE 1445

Qy 351 -----TFQNDKG---BAIDKDNGLVTDSSKLVFLGRGDKKEYTGDFEDKPNVIAKEDGSM 402  
Db 1446 ATSELTLQDQNQSLSTEIEKTAALKTSSK-----DLEVCGNOKSLO-----DSLK 1493

QY 403 PTDTPKPVLSPMDKPNVYENPSKANPKIVYRNPEFY----LRGKISDKOGFPNWLRLRVNESWVD 457  
Db 1494 SVKSELENKFENKYNQETTSLSKAKIEEKQETIVTQLQTELKDRISERVERAMISENSETV- 1552  
QY 458 NYLIYGDHLHINTRDENIKUNVKDGDIMDWMGKDYKANGFDPKVUTMDGNVYLQGSDL 517  
Db 1553 -----KEYS-----SKI 1567  
QY 518 NAKAVGVHYQFLDNVKPENVPKGNTSIEY----ADGSKSVNINDRNGNDFGEI 571  
Db 1568 NS-----I-KENHSKEITTHNEQKTSLSKDIAKLSQDIESAQTOLEDGENQNLKLA 1618  
QY 572 Q-EQHTYINGKEYTSF----NDIKO1D--KTYLNKLUVKOPFARNNTVKEB---ILNKD 620  
Db 1619 SLEKH--NTPSATSLERKNMOKIELSETIMSLTKBLTGSDAKOSQKEYKLTKNSD 1675  
QY 621 T----GEVSUPLKPHRTVTQNGKEMSTTVESEEPILPYKGELEGKIQEDGMEISGF 675  
Db 1676 TE SKLEROLELBE--KVKSDLQTADEKUGJTERETAL---KSELET----- 1717  
QY 676 ESGKKDAGVYVNLKDPIKPKKIERKESENKPPDVSKKDNQVNHSQLNESHRK- 734  
Db 1718 --VKNSGLISTTSELAAALKTW-KSLEKEKSLQFELSGNKSTLEDYIQKHSIDSEKIAL 1774  
QY 735 EDLORENSHSQSDSTK-----DVTAVLDKNNISSKSTTNPNK 773  
Db 1775 TDELKEKTKQDDSKKKLTETEJENDLIST--KKELETEKTOTSKFK 1817

**RESULT 14**  
US-01-742-096-3  
; Sequence 3, Application US/09742096  
; Patent No. US20020155441A1  
; GENERAL INFORMATION:  
; APPLICANT: DRUILHE, PIERRE  
; APPLICANT: DAUBERSIES, PIERRE  
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
; FILE REFERENCE: 200773US0D1V  
; CURRENT APPLICATION NUMBER: US/09/742, 096  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 08/973, 642  
; PRIOR FILING DATE: 1998-02-06  
; PRIOR APPLICATION NUMBER: PCT/FR96/00894  
; PRIOR FILING DATE: 1996-06-12  
; PRIOR APPLICATION NUMBER: FR 95/07007  
; PRIOR FILING DATE: 1995-06-13  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 1786  
; TYPE: PRT  
; ORGANISM: *P. falciparum*  
; US-09-742-096-3

Query Match 5.0%; Score 200.5; DB 9; Length 1786;  
Best Local Similarity 20.5%; Pred. No. 0.015; Gaps 52;  
Matches 204; Conservative 160; Mismatches 329; Index 301; Gaps 52;

QY 6 AES-KPKN-1GNGKKGSKLKD-----TGVERHHQNE-----ESTK 40  
Db 868 AESVTVPNNEBLLQENTITDIEKLEHENVLSAULENTOSBERKEVIDVEVK 927  
QY 41 EKSFETI-----DRNISTRD-FEN-----KDLKLUKKPFREYDFTS 78  
Db 928 BEVATTIETWQAEERKSANTITEIPELNLENAEVENVENLEKLNETVTFNLKVE 987  
QY 79 ET--GKRM--EYKY----DDKG--NLYDDGSDLEVT-----EKLD-- 114  
Db 988 EWEVTSGESLHNENMKAFFSBIFDNVKGQIBDNLT--GMRSTSTIVQSEEKVDLN 1044  
QY 115 -EIKSKYGVYSPSKOCHFETLGKLSNVSKAKV-----YYGNVYKSTIKATKDYFH 166  
Db 1045 ENWVSSLDNTENMKG--LINKLENISSESBGVTWTEVEONVWVWDVPMK--- 1097

**RESULT 15**  
US-10-415-253-2  
; Sequence 2, Application US/10415253  
; Publication No. US20040067236A1  
; GENERAL INFORMATION:  
; APPLICANT: Cohen, Joe  
; APPLICANT: Drulhe, Pierre  
; TITLE OF INVENTION: Immunogenic Compositions Comprising  
; TITLE OF INVENTION: Liver Stage Malarial Antigens  
; FILE REFERENCE: B45250  
; CURRENT APPLICATION NUMBER: US/10/415, 253  
; CURRENT FILING DATE: 2003-04-25  
; PRIOR APPLICATION NUMBER: PCT/EP01/12349  
; PRIOR FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: EP00203724.0  
; PRIOR FILING DATE: 2000-10-25  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 2

QY 167 SKMTFUDIYANINDIVGLAFAGDMRLFVKNDQKAKIKIRMPKEKIKTSEPYPVSSY 226  
Db 1098 -----DOFGIGLNBAGGL----KEMPFENLDEDVRESBESPVITVERBKDGPVQKEVEKET 1146  
QY 227 GNVIELEGGD----LSKNCPDNLTM-----ESGKYYDSBKRQQVLLKDNILRKGYALK 277  
Db 1147 VSIEEMMENIVDVLBEEKSDLTDMIAVEESIISDSKEETESIKDK--EKDSLV 1203  
QY 278 VTTYNPG-----KTDMLEGNGVSK--EDIAKIOKANPL----- 310  
Db 1204 VEEQDNMDDESVEKULKOMBEILMDAVEINIDTSKJETEBLINEFADLKIDMEK 1263  
QY 311 -----RALSETTIYADSRNVEDGR--STQSULMSALDGFIIRYQVFTRKNDKGBAIDK 363  
Db 1264 LKELEKLHSE----DSKEIIDAODDTLEKVEEHD-----ITTLDEVELKDV 1309  
QY 364 DGNLVTDSSKVLFLGKDKDEYGEDKPNVIAKEDGSMFLRIDTPVNL-----SM 413  
Db 1310 EBDKLEKVSDL---KDELEEDILK---VCEIKELESELEDEYKELKTIEDILEBKEI 1362  
QY 414 DKNYNPFSKSNKIVYRNPEFYLRGKTSKDGCFNWLRLVRNSWVD-----NYLIYGD 465  
Db 1363 EKDHFEKPFEEBAEIKOLEADILKEV-----SLEVEBEKKLEFHLCBVEVHITGDA 1417  
Db 1418 HIKGLBEBDDBE-BVDDLGSTSILDMLKGMELGDMKESLBDVTTKGERV---ESIKDVL 1473  
QY 519 AKAVGVHYQFLYDN-----VPEVNIDKGNTSIEYADCKSVFNINIK--RN 564  
Db 1474 SSALGMNDBEQMTRKAQRPKLEEVILKESVKEEKKKIT-----KKVQFDPDOKKEPK 1528  
QY 565 NGFDGETQEOHITYNGK-----EVTSFNDIKQIDKTIN-----IKTVKDFAR--N 609  
Db 1529 EIVEVNMKDEIEDVEDDEIDEDBEDKVEDIDEIDEDGEDKOBVIDIIVQKREK 1588  
QY 610 TTVKERFLINKDGTG--VSELKPH-----RVTUTONGKEMSTTVESEED 651  
Db 1589 VKAKKKKLEKKVEEGVSGLKRKRHVDEVWKVYQKIDKEVDEKVSALLEKSNDDVNLVQKQNQD 1648  
QY 652 FILPVYKGELEKQYQFDGWESFGFGKQDAGYVIM-----SKDTFKPKVFK 698  
Db 1649 FFSKV-KNFVCKYVFAFPFSAVA--FASYVVFRTSFLFSSCVTASSTYLISKVFK 1705  
QY 699 KIEKEKEENKP----TDFV-----SKKKNPQVNHSQLNESHRKDQLR 739  
Db 1706 TINKKE--RPFYTFVDFDIFKLNKHYLQQMKEKSKSEKONNIVBT--NKAEGKGNVQ 1760  
QY 740 EBSOKSDSTKDVTAVLDKNNISSKSTTNPNK 773  
Db 1761 TNKTEKT-----TKVDKNNKVKPKKRTQSK 1786

; LENGTH: 1787  
; TYPE: PRT  
; ORGANISM: K1 Parasite Clonee  
US-10-415-253-2

; LENGTH: 1787  
; TYPE: PRT  
; ORGANISM: K1 Parasite Clone  
US-10-4155-253-2

OY	740	EEHSQSDSTKDVTATVLDKNNISKSSTTNPNK	773
Db	1761	TNKTEKT-----TKDQNKVKPKRRTQSK	1786

Query Match Similarity: 5.0%; Score 200.5; DB 15; Length 1787; Best local

Search completed: February 16, 2005, 19:32:38  
Search time : 147 secs

QY 6 AES--KFKN-LGMCKEGSLKKOT-----ESTK 40  
 Db 868 AEEVTTSNILEIQTQNTINDTIEKLELHWLVAALENTOSBEEKGEVIVIEUK 927  
 QY 41 EKKSFTI-----DRNSTIRD-FEN-----KOLKLIKKEFREVDDETS 78  
 Db 928 BEVATTLIBTVEQEKEKSANTITIFENLEENAVENNAVENLKLNETVENTVLDKVE 987  
 QY 79 ET---GKRM- BYDVK-----DDKG---NIAYDDTLEYT-----BKLD--- 114  
 Db 988 ETWEISGSESLENNEMDKAPFSFELDNVKIQEMLT---GMFRESITSIVIQSERKVUDL 1044  
 QY 115 -ETRSKTYGVLSPLSPSKDHPEILGKISNYSKNAKY-----YCQNNYKSIETKATKYDFH 166  
 Db 1045 ENVSILSDNINNMKEG---LUNKLEN-SSTEGQVETTEHVEONVYDVDPAMK--- 1097  
 QY 167 SKMNTFDLYANTINDIVDGLAFAGDMRLFYKDNDOKKAECIRMPKKEKETKSEVPPVSSY 226  
 Db 1098 -----DOPFLGILNEAGGL-----KEMFNFLLEDVFVKSESDBVITVEIKOPVOKVERET 1146  
 QY 227 GNTIELGSD-----LSKONPDNLTM-----ESKIIYSDESKOQLLKONILRGYALK 277  
 Db 1147 VSLEMEMENTIVOLEBEKEKDUDKMDIAVEESTEISSUSKEESIOK---EKDVSLV 1203  
 QY 278 VTVTYPNG-----KUTMLEGENGVSK---EDTAKIOKANPNT----- 310  
 Db 1204 VEEVONDMDSESVKVELKNMREBLMKDAVEINIDTSKLIKEQBLENNEVAIDLKDMEK 1263  
 QY 311 -----RALSETTIYADSRNVEDGR---STOSVMSALDGENDIIRYQVFPTKONDGEAIDK 363  
 Db 1264 LKELEKALSE-----DSKBTTIDAOKDTLEKVIEBHD-----ITTLDEVELKOV 1309  
 QY 364 DGNLVTDSKKULVFGKDDKEYTGDKKFNVTEAKEDGSMFPIDTKVNLT-----SM 413  
 Db 1310 EERKIKYDSL-----KOLBEDILKE-----VKGKALESELVEDYKELKTIEDILEKKEI 1362  
 QY 414 DKONYFPNSKSNKTYVRNPFPYLRKISDKGGFNLHMLRVNESWVD-----NYLIGDL 465  
 Db 1363 EKORFEKEBEEABERKOLRADILKEVS-----SIEVEERCKLEWHELKEBEVENIISGD 1417  
 QY 466 HIDNTRDENKLNKWD-GDIM---DNMGKDYKANGFPDKVTDMDGAVLQTCYSDLN 518  
 Db 1418 HIKGLEEDDLE-EVTDLKGSISLDNMKGDMWLGMDMKESLEDVUTTKLGERV---ESTKDV 1473  
 QY 519 AKAVGWHQFLYDN-----VKUEVNIDPKGNTSIEYADGKSVWNINDK-RN 564  
 Db 1474 SSALGMDBEQMKTKKAQRPKLEEVULKKEVKEPKEKKIT-----KKYRFDIKOKEPD 1528  
 QY 555 NGFDGEBOHOHYNGK-----EYTSENDIKOIDKTLN-----IKUWKDFPAR-N 609  
 Db 1529 EIVEVMKEDIEDEVEDJEEDIEBEDKVEDIDIDEDEDGEDKEDKEDVOLIVOKERKIEK 1588  
 QY 610 TTWKEPLINKDTGE-VSELKPH-----RWTWTOQNGKEMSTSIVSEED 651  
 Db 1589 VRKAKKKKLBKVKVEGGVSGLKHVKDUMKTYVQKIDKEVSKALESKUDVNTLKQND 1648  
 QY 652 FILPVYKGBLEKGFQDFGHWBISGEGKKGODAGYVNL-----SKDTKIPVK 698  
 Db 1649 FFSKV-KNFTKVKYFAAPPISAVA- FASIVVGPFTSLFSSCUTIASSTYKISKVD 1705  
 QY 699 KIBERKEEENKP-----TEFW-----SKKDNPOHNSOLESHEKEDLOR 739  
 Db 1706 TINKKE---RPFVSVFDFIKNTHQLOOMKEKEKSKEKONNNFBT--NKAEGKGNVQ 1760

This Page Blank (uspto)

Scoring table: BLOSUM62 Gapop 10.0 , Gapext 0.5									
Searched: 1612378 seqs, 512079187 residues									
Total number of hits satisfying chosen parameters: 1612378									
Minimum DB seq length: 0									
Maximum DB seq length: 2000000000									
Post-processing: Minimum Match 0% Maximum Match 100% Listing first 45 summaries									
Database : UniProt 03::*									
1: uniprot_sprot:*									
2: uniprot_trembl:*									
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.									
SUMMARIES									
Result No.	Score	Query Match Length	DB ID	Description					
1	4026	100.0	2119	2 Q9AHT5	Q9AHT5	PRELIMINARY;	PRT;	2119 AA.	RESULT 1
2	4026	100.0	2140	2 Q9TRX6	Q9AHT5;				RQ
3	3793.5	94.2	2144	2 Q9SAM8	Q9AHT5;				RC
4	3709.5	92.1	2144	2 Q8DQP7	Q9AHT5;				RA
5	270.5	6.7	2649	2 Q7RAS7	Q9AHT5;				DT
6	265.5	6.6	1642	2 Q8IBB4	Q9AHT5;				DT
7	264	6.6	3504	2 Q8IL45	Q9AHT5;				01-JUN-2001 (TrEMBLrel. 17, last sequence update)
8	259	6.5	1850	2 Q7RGQ2	Q9AHT5;				01-MAR-2004 (TrEMBLrel. 26, last annotation update)
9	259	6.4	2157	2 Q7RRG9	Q9AHT5;				DE
10	258.5	6.4	1811	2 Q7REB9	Q9AHT5;				Serine protease (Fragment).
11	257.5	6.4	2661	2 Q7RMS4	Q9AHT5;				GN
12	257	6.4	1389	2 Q7RQJ4	Q9AHT5;				OS
13	255	6.3	1777	2 Q8IP34	Q9AHT5;				OC
14	253.5	6.3	1127	2 Q8YV76	Q9AHT5;				Streptococcus pneumoniae; Streptococcaceae;
15	252.5	6.3	2227	2 Q8I121	Q9AHT5;				NCBI_TaxID=1313;
16	252.5	6.3	2723	2 Q7RQB6	Q9AHT5;				RN
17	251.5	6.2	2849	2 Q8HY4	Q9AHT5;				[1]
18	251.5	6.2	3391	2 Q8I2V4	Q9AHT5;				SEQUENCE FROM N.A.
19	251.5	6.2	3519	2 Q8IE65	Q9AHT5;				RESTRAIN4; MEDLINE=21116975; PubMed=1117332;
20	250	6.2	1674	2 Q8ILU2	Q9AHT5;				DOI=10.1128/IAI.69.3.1591-1598.2001;
21	250	6.2	1650	2 Q773Z8	Q9AHT5;				Wizemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C., Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomola E., Gayle A., Brewen Y.A., Walsh W., Barron P., Lathigra R., Hanson M., Langemann S., Johnson S.N., Koenig S.,
22	247.5	6.1	5767	2 Q8IS25	Q9AHT5;				RT "Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection.";
23	246.5	6.1	2033	2 Q8TM18	Q9AHT5;				-SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (By similarity).
24	245	6.1	2269	2 Q8ITIA2	Q9AHT5;				CC EMBL; AF291699; AKI19159.1; -.
25	242.5	6.0	2664	2 Q7RL0	Q9AHT5;				DR HSSP; P00782; 28BT.
26	242.5	6.0	3317	2 Q8EW8B	Q9AHT5;				DR MEROPS; S08.064; -.
27	241.5	6.0	1033	2 Q8IBBB	Q9AHT5;				DR GO; GO_009986; C:cell surface; IEA.
28	241.5	6.0	2694	2 Q7RQJ1	Q9AHT5;				DR GO; GO_0005618; C:cell wall; IEA.
29	241.5	6.0	3063	2 Q61MC1	Q9AHT5;				DR GO; GO_000233; F:peptidase activity; IEA.
30	240.5	6.0	2740	2 Q7RF52	Q9AHT5;				DR GO; GO_00425; :subtilase activity; IEA.
31	240.5	6.0	4433	2 Q8IJU5	Q9AHT5;				DR GO; GO_0005020; P:proteolytic and peptidolysis; IEA.
32	240	6.0	1455	2 Q8IKG8	Q9AHT5;				DR IPR01435; DUF1034.
33	240	6.0	10661	2 Q8I3Z1	Q9AHT5;				DR InterPro; IPR0189; Gram_pos_anchor.
34	239.5	5.9	2569	2 Q8IBG8	Q9AHT5;				DR InterPro; IPR00137; PA.
35	239	5.9	2586	2 Q7PDT7	Q9AHT5;				DR InterPro; IPR00209; Pept_S8_S53.
36	238	5.9	1081	2 Q8IXL2	Q9AHT5;				DR InterPro; IPR01259; Prot_im_S8A.
37	238	5.9	3322	2 Q8IKL0	Q9AHT5;				DR InterPro; IPR001680; WD40.
38	238	5.9	3628	2 Q86BY0	Q9AHT5;				DR pfam; PF06280; DUF1034; 1.
39	238	5.9	3704	2 Q8IRK8	Q9AHT5;				DR pfam; PF0746; Gram_pos_anchor.
40	237.5	5.9	1104	2 Q7RSQ8	Q9AHT5;				DR pfam; PF0225; PA; 1.
41	237.5	5.9	3535	2 Q8IC29	Q9AHT5;				DR pfam; PF0522; Peptidase_S8; 1.
42	237	5.9	2511	2 Q8II44	Q9AHT5;				DR pfam; PF00723; Subtilisin.
43	237	5.9	5229	2 Q7RTF4	Q9AHT5;				DR pfam; PF06280; DUF1034; 1.
44	236.5	5.9	1114	2 Q87242	Q9AHT5;				DR pfam; PF0746; Gram_pos_anchor.
45	236.5	5.9	1738	2 Q8ITAL5	Q9AHT5;				DR pfam; PF06280; DUF1034; 1.

Q8ikg8 plasmidium  
Q8i3z1 plasmidium  
Q8ibg8 plasmidium  
Q7pd7 plasmidium  
Q8ixl2 clostridium  
Q8ikl0 plasmidium  
Q96by0 plasmidium  
Q8irk8 plasmidium  
Q7rsq8 plasmidium  
Q8ic29 plasmidium  
Q8il44 plasmidium  
Q97242 plasmidium  
Q8ials plasmidium

## ALIGNMENTS

RESULT 1  
 Q9AHT5 PRELIMINARY; PRT; 2119 AA.  
 ID Q9AHT5;  
 AC Q9AHT5;  
 DT 01-JUN-2001 (TrEMBLrel. 17, last sequence update)  
 DR 01-MAR-2004 (TrEMBLrel. 26, last annotation update)  
 DE Serine protease (Fragment).  
 GN Name=ppta;  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 STREPTOCOCCUS pneumoniae; Streptococcaceae;  
 RX DOI=10.1128/IAI.69.3.1591-1598.2001;  
 RA Infec. Immun. 69:1593-1598 (2001);  
 CC -SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (By similarity).  
 DR EMBL; AF291699; AKI19159.1; -.  
 DR HSSP; P00782; 28BT.  
 DR GO; GO\_009986; C:cell surface; IEA.  
 DR GO; GO\_0005618; C:cell wall; IEA.  
 DR GO; GO\_000233; F:peptidase activity; IEA.  
 DR GO; GO\_00425; :subtilase activity; IEA.  
 DR GO; GO\_0005020; P:proteolytic and peptidolysis; IEA.  
 DR InterPro; IPR01435; DUF1034.  
 DR InterPro; IPR0189; Gram\_pos\_anchor.  
 DR InterPro; IPR00137; PA.  
 DR InterPro; IPR00209; Pept\_S8\_S53.  
 DR InterPro; IPR01259; Prot\_im\_S8A.  
 DR InterPro; IPR001680; WD40.  
 DR pfam; PF06280; DUF1034; 1.  
 DR pfam; PF0746; Gram\_pos\_anchor.  
 DR pfam; PF0225; PA; 1.  
 DR pfam; PF0082; Peptidase\_S8; 1.  
 DR pfam; PF0522; Subtilisin; 1.  
 DR pfam; PF00723; Subtilisin.  
 DR pfam; PF06280; DUF1034; 1.  
 DR PROSITE; PS05047; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE; PS0137; SUBTILASE\_HIS; UNKNOWN 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN 1.  
 DR PROSITE; PS00679; WD\_REPEATS\_1; UNKNOWN 1.  
 DR KW Cell wall; Peptidoglycan-anchor; Protease.  
 FT NON\_TER 1

SEQ	SEQUENCE	2119 AA;	238227 MW;	517F997F6B66A6A CRC64;	RP
Query Match	100.0%;	Score 4026;	DB 2;	Length 2119;	SEQUENCE FROM N-A.
Best Local Similarity	100.0%;	Pred. No. 2e-161;	Length	Indels 0;	STRAIN=ATCC BAA-334 / TIGRA;
Matches	773;	Conservative	0;	Gaps 0;	MEDLINE=21157209; PubMed=11463916; DOI=10.1126/science.1061217;
QY	1 KLGIAESFKENKGKEGSILKDTGVEHHOGENESEIKKSFTIRNISTRDFENK	60	DB	Teteelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,	
QY	1313 KLGEIAESFKENKGKEGSILKDTGVEHHOGENESEIKKSFTIRNISTRDFENK	60	DB	Peterson S.N., Heidelberg J.F., Deboy R.R., Haft D.H., Dodson R.J.,	
QY	61 DLKLUKKRKFREDDFTSETGKRMEMBYDYKDDKGNIAYDDGDLTEYEKEDEIKSKI	120	DB	Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,	
QY	1373 DLKLUKKRKFREDDFTSETGKRMEMBYDYKDDKGNIAYDDGDLTEYEKEDEIKSKI	1432	DB	Umayam L.A., White O., Salzberg S.L., Lewis M.R., Redune D.,	
QY	121 YGVLSPSKOGHFEITIGKISNVSKNAKVYGNNSKIEIKATKDYHISKMTFDLYANIND	180	DB	Rothzapfel E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,	
QY	1433 YGVLSPSKOGHFEITIGKISNVSKNAKVYGNNSKIEIKATKDYHISKMTFDLYANIND	1492	DB	Holtzapfel L.A., Feldblum T.V., Angiolillo S.V., Dickinson T.,	
QY	181 IVDGLAFAGDMRLFKNDQKAEIKIRMPBKETKSEPYVSYGNVIELGEGDLSK	240	DB	Hickey B.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,	
QY	1493 IVDGLAFAGDMRLFKNDQKAEIKIRMPBKETKSEPYVSYGNVIELGEGDLSK	1552	DB	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;	
QY	241 KPDNLTKMEGKISDSERQQLIYKDNITLRKGALKVTYNGKTDMLRGNGVTSKEDI	300	DB	"Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.", 293:498-506 (2001).	
QY	1553 KPDNLTKMEGKISDSERQQLIYKDNITLRKGALKVTYNGKTDMLRGNGVTSKEDI	1612	DB	Science 293:498-506 (2001).	
QY	301 AKTOKANPNURALSETTYIADSRSNVEDGRSNTOSVMSALDGFTNTRYQVFTRKMDNGEA	360	DB	-1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (By similarity).	
QY	1613 AKTOKANPNURALSETTYIADSRSNVEDGRSNTOSVMSALDGFTNTRYQVFTRKMDNGEA	1672	DB	an amide bond (By similarity).	
QY	361 IDKQNLVTDSSKLULFGKDKKEYGEDKVNKAKEGDGMLFLTPKPVNLSMDQKHFNP	420	DB	DR; GO:000508; P:proteolysis and peptidolysis; IEA.	
QY	1673 IDKQNLVTDSSKLULFGKDKKEYGEDKVNKAKEGDGMLFLTPKPVNLSMDQKHFNP	1732	DB	DR; GO:000508; P:proteolysis and peptidolysis; IEA.	
Db	421 SKSNKIYVRNPEFYRKISDSKGHHWNLRLVNESTVDNYLITYGDAIDNTRDFNITLNK	480	DR	DR; GO:000508; P:proteolysis and peptidolysis; IEA.	
QY	1733 SKSNKIYVRNPEFYRKISDSKGHHWNLRLVNESTVDNYLITYGDAIDNTRDFNITLNK	1792	DR	DR; GO:000508; P:proteolysis and peptidolysis; IEA.	
QY	481 DGDIMDWGMKDYKANGFPDKUTDMGNGVYLQTYGDLNAKAVGHTYQFLDNVKPVNID	540	DR	DR; InterPro; IPR01435; DUF1034.	
Db	1793 DGDIMDWGMKDYKANGFPDKUTDMGNGVYLQTYGDLNAKAVGHTYQFLDNVKPVNID	1852	DR	DR; InterPro; IPR01435; DUF1034.	
QY	541 PKGTSIEVADGKSVFNLNDKRNQGFDGSIQBOQHYYINGKEYTFSPNDIKOIIDYLNK	600	DR	DR; InterPro; IPR00209; Pept_S8_S53.	
Db	1853 PKGTSIEVADGKSVFNLNDKRNQGFDGSIQBOQHYYINGKEYTFSPNDIKOIIDYLNK	1912	DR	DR; InterPro; IPR010259; Prot_inh_S8A.	
QY	601 IVVKDPARNITVKERILNKOTGEVSLKPHRTVUTONGKEMSTIVSERDPLPYKGE	660	DR	DR; InterPro; IPR00723; Subtilisin.	
Db	1913 IVVKDPARNITVKERILNKOTGEVSLKPHRTVUTONGKEMSTIVSERDPLPYKGE	1972	DR	DR; InterPro; IPR00723; Subtilisin.	
QY	661 LEKGYQFDWEISGREGKODAGYVYNLSKOTFKVFKTEEKERENKPTFDVSKKDN	720	DR	DR; PROSITE; PS050847; Gram_Pos_Anchor; 1.	
Db	1973 LEKGYQFDWEISGREGKODAGYVYNLSKOTFKVFKTEEKERENKPTFDVSKKDN	2032	DR	DR; PROSITE; PS050847; Gram_Pos_Anchor; 1.	
QY	721 PQVNLSQLNEHRKDLOREBHSQSSTDKVATVLDKRNISSTTNPNK	773	DR	DR; PROSITE; PS05138; Subtilase_Ser; UNKNOWN_1.	
Db	2033 PQVNLSQLNEHRKDLOREBHSQSSTDKVATVLDKRNISSTTNPNK	2085	DR	DR; PROSITE; PS050678; WD_Repeats_1; UNKNOWN_1.	
RESULT 2			DR	KW Cell wall; Complete proteome; Peptidoglycan-anchor; Protease.	
Q97RY6	PRELIMINARY; PRT; 2140 AA.		DR	SEQUENCE 2140 AA; 240426 MW; FA44AB82938B334 CRC64;	
AC	Q97RY6; 01-OCT-2001 (Tremblrel. 18, Created)		DR	DR; PROSITE; PS050847; Gram_Pos_Anchor; 1.	
DT	01-OCT-2001 (Tremblrel. 18, Last sequence update)		DR	DR; PROSITE; PS050847; Gram_Pos_Anchor; 1.	
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)		DR	DR; PROSITE; PS050847; Gram_Pos_Anchor; 1.	
DE	Serine protease, subtilase family.		DR	DR; PROSITE; PS050847; Gram_Pos_Anchor; 1.	
GN	Orderedsocusnames-SP0641;		DR	DR; PROSITE; PS050847; Gram_Pos_Anchor; 1.	
OS	Streptococcus pneumoniae.		DR	DR; PROSITE; PS050847; Gram_Pos_Anchor; 1.	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		DR	DR; PROSITE; PS050847; Gram_Pos_Anchor; 1.	
NCBI_TaxID=1313;			DR	DR; PROSITE; PS050847; Gram_Pos_Anchor; 1.	
RN			DR	DR; PROSITE; PS050847; Gram_Pos_Anchor; 1.	
QY	1 KGEIAESFKENKGKEGSILKDTGVEHHOGENESEIKKSFTIRNISTRDFENK	60	DB	1334 KLGIAESFKENKGKEGSILKDTGVEHHOGENESEIKKSFTIRNISTRDFENK	
QY	61 DLKLUKKRKFREDDFTSETGKRMEMBYDYKDDKGNIAYDDGDLTEYEKEDEIKSKI	120	DB	QY 61 DLKLUKKRKFREDDFTSETGKRMEMBYDYKDDKGNIAYDDGDLTEYEKEDEIKSKI	
QY	1394 DLKLUKKRKFREDDFTSETGKRMEMBYDYKDDKGNIAYDDGDLTEYEKEDEIKSKI	1453	DB	QY 1394 DLKLUKKRKFREDDFTSETGKRMEMBYDYKDDKGNIAYDDGDLTEYEKEDEIKSKI	
QY	121 YGVLSPSKOGHFEITIGKISNVSKNAKVYGNNSKIEIKATKDYHISKMTFDLYANIND	180	DB	QY 121 YGVLSPSKOGHFEITIGKISNVSKNAKVYGNNSKIEIKATKDYHISKMTFDLYANIND	
QY	1454 YGVLSPSKOGHFEITIGKISNVSKNAKVYGNNSKIEIKATKDYHISKMTFDLYANIND	1513	DB	QY 1454 YGVLSPSKOGHFEITIGKISNVSKNAKVYGNNSKIEIKATKDYHISKMTFDLYANIND	
QY	181 IVDGLAFAGDMRLFKNDQKAEIKIRMPBKETKSEPYVSYGNVIELGEGDLSK	240	DB	QY 181 IVDGLAFAGDMRLFKNDQKAEIKIRMPBKETKSEPYVSYGNVIELGEGDLSK	
Db	1514 IVDGLAFAGDMRLFKNDQKAEIKIRMPBKETKSEPYVSYGNVIELGEGDLSK	1573	DB	Db 1514 IVDGLAFAGDMRLFKNDQKAEIKIRMPBKETKSEPYVSYGNVIELGEGDLSK	
QY	241 KPDNLTKMEGKISDSERQQLIYKDNITLRKGALKVTYNGKTDMLRGNGVTSKEDI	300	DB	QY 241 KPDNLTKMEGKISDSERQQLIYKDNITLRKGALKVTYNGKTDMLRGNGVTSKEDI	
QY	1574 KPDNLTKMEGKISDSERQQLIYKDNITLRKGALKVTYNGKTDMLRGNGVTSKEDI	1633	DB	QY 1574 KPDNLTKMEGKISDSERQQLIYKDNITLRKGALKVTYNGKTDMLRGNGVTSKEDI	
QY	301 AKTOKANPNURALSETTYIADSRSNVEDGRSNTOSVMSALDGFTNTRDFNITLNK	1693	DB	QY 301 AKTOKANPNURALSETTYIADSRSNVEDGRSNTOSVMSALDGFTNTRDFNITLNK	

QY	361	IKDGNVLTDSKLVIFGKDDKEYTGDKFNEAKEDGSMFLIDTPKVNLISMDKOVFNP	420	DR PRINTS; PRO0723; SUBTILLISIN.
Db	1694	IKDGNVLTDSKLVIFGKDDKEYTGDKFNEAKEDGSMFLIDTPKVNLISMDKOVFNP	1753	DR TIGRMS; TIGR01167; LPXIG anchor; 1.
QY	421	SRSNKTYVRNPFYLRKISDKGGFWNLRVNESVVDNYLYGDLHIDNTDFNKLNVK	480	DR POSITIVE; PS50847; GRAM POS ANCHORING; 1.
Db	1754	SRSNKTYVRNPFYLRKISDKGGFWNLRVNESVVDNYLYGDLHIDNTDFNKLNVK	1813	DR PROSITE; PS00137; SUBTILASE HIS; UNKNOWN 1.
QY	481	DODIMPGMKOYKANGFPDKYTDGMVNLYQGSDLNAKAVGHYQFLYKVKPEVNID	540	DR PROSITE; PS00118; SUBTILASE SER; UNKNOWN 1.
Db	1814	DODIMPGMKOYKANGFPDKYTDGMVNLYQGSDLNAKAVGHYQFLYKVKPEVNID	1873	DR PROSITE; PS00678; WD REPEATS 1; UNKNOWN 1.
QY	541	PKGNTSIEYDGSKVVENINDKNGDFGEQEHYINGKEYTSNDIKOIDKLNIK	600	KW Cell wall; Peptidoglycan-anchor; Signal.
Db	1874	PKGNTSIEYDGSKVVENINDKNGDFGEQEHYINGKEYTSNDIKOIDKLNIK	1933	FT SIGNAL_1 Potential.
QY	601	IVVKDFARNTTKEFIANKDGTGEVSELKPHRTVTONGKEMSTVSEEDFILPYKGE	660	FT CHAIN 20 2144 PRTA.
Db	1934	IVVKDFARNTTKEFIANKDGTGEVSELKPHRTVTONGKEMSTVSEEDFILPYKGE	1993	FT SEQUNCE 2144 AA; 240725 MW; 2052511470/41131 CRC64;
QY	661	LEKGYOPDGWEISGFECKKDAGYVINSKDKTIPKFKKEKKERKEENKPKFDVSKDN	720	Query Match 94.2%; Score 3793.5; DB 2; Length 2144;
Db	1994	LEKGYOPDGWEISGFECKKDAGYVINSKDKTIPKFKKEKKERKEENKPKFDVSKDN	2053	Best Local Similarity 94.7%; Pred. No. 1.3e-151; Indels 1; Gaps 1;
RESULT 3				Matches 732; Conservative 16; Mismatches 1;
Q9S4MB		PRELIMINARY; PRT; 2144 AA.		Db 1339 KLGRIEBSKFKQLNKNTDSLNKETAEVENNLVUNQNSIECKSLFENHKITSTIREFNK 60
ID	Q9S4MB			Qy 61 DDKKLKRRKFREVDFFSETGRMRMERYDYKDKGNIITYDGTDLDEYETKLDEKSKI 120
AC				Db 1399 DDKKLKRRKFQBDFFN-GGTRTVERDYKYDKDGNITAYDGTDLDEYETKLDEKSKI 1457
DT	01-MAY-2000 (T-BMBLcl. 13, Created)			Qy 121 YCVLSPSKDGHFELLKISNSVSKAQYQGNNYKSTIEKATKDYFHSKTWTDFLYANIND 180
DT	01-MAY-2000 (T-BMBLcl. 13, Last sequence update)			Db 1458 YCVLSPSKDGHFELLKISNSVSKAQYQGNNYKSTIEKATKDYFHSKTWTDFLYANIND 1517
DR	01-MAR-2004 (T-BMBLrel. 26, Last annotation update)			Qy 181 IVGLABAGDMRLFVKUNDQKAEEIKRMPKKEKTEYKVVSYGAVBEGEGLSKN 240
GN	NameprtA;			Db 1518 IVGLAFAGDMLFVKUNDQKAEEIKRMPKKEKTEYKVVSYGAVBEGEGLSKN 1577
OS	Streptococcus pneumoniae.			Qy 241 KPDNLTMESGKISYDSEKQYLKONTILRGYALKVTNPKGTDMLLEGNGVSKEDI 300
OC	Bacteria; Firmicutes; Lactobacillales; Streptococaceae; Streptococcus.			Db 1578 KPDNLTMESGKISYDSEKQYLKONTILRGYALKVTNPKGTDMLLEGNGVSKEDI 1637
OX	NCBI_TaxID=1313;			Qy 301 AKIOKAPNPNLALSETTYADSRNVEGDSRSTSOLMSALDFNIRYQVFPMKNDGKA 360
RN	[1]			Db 1638 AKIOKAPNPNLALSETTYADSRNVEGDSRSTSOLMSALDFNIRYQVFPMKNDGKA 1697
RP	SEQUENCE FROM N.A.			Qy 361 IKDGNVLTDSKLVIFGKDDKEYTGDKFNEAKEDGSMFLIDTPKVNLISMDKOVFNP 420
RC	STRAIN=3.B;			Db 1698 IKDGNVLTDSKLVIFGKDDKEYTGDKFNEAKEDGSMFLIDTPKVNLISMDKOVFNP 1757
RX	MEDLINE=21585565; PubMed=11728722;			Qy 421 SRSNKTYVRNPFYLRKISDKGGFWNLRVNESVVDNYLYGDLHIDNTDFNKLNVK 480
RA	Bethe G., Nau R., Wellmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,			Db 1758 SRSNKTYVRNPFYLRKISDKGGFWNLRVNESVVDNYLYGDLHIDNTDFNKLNVK 1817
RA	Zybk G.;			Qy 481 DODIMPGMKOYKANGFPDKYTDGMVNLYQGSDLNAKAVGHYQFLYKVKPEVNID 540
RT	"The cell wall-associated serine protease Prta: a highly conserved virulence factor of <i>Streptococcus pneumoniae</i> ."			Db 1818 DODIMPGMKOYKANGFPDKYTDGMVNLYQGSDLNAKAVGHYQFLYKVKPEVNID 1877
RL	FEMS Microbiol Lett. 205:99-104 (2001).			Qy 541 PKGNTSIEYDGSKVVENINDKNGDFGEQEHYINGKEYTSNDIKOIDKLNIK 600
CC	- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (BY SIMILARITY).			Db 1878 PKGNTSIEYDGSKVVENINDKNGDFGEQEHYINGKEYTSNDIKOIDKLNIK 1937
CC	EMBL: AF17143; AACD8399.1; -.			Qy 601 IVVKDFARNTTKEFIANKDGTGEVSELKPHRTVTONGKEMSTVSEEDFILPYKGE 660
DR	HSS; P00782; 2SBT.			Db 1938 IVVKDFARNTTKEFIANKDGTGEVSELKPHRTVTONGKEMSTVSEEDFILPYKGE 1997
DR	MERO; S08_064; -.			Qy 661 LEKGYOPDGWEISGFECKKDAGYVINSKDKTIPKFKKEKKERKEENKPKFDVSKDN 720
DR	GO: GO:0009986; C:cell surface; IEA.			Db 1998 LEKGYOPDGWEISGFECKKDAGYVINSKDKTIPKFKKEKKERKEENKPKFDVSKDN 2057
DR	GO; GO:00561; C:cell wall; IEA.			Qy 721 PQVNHSQINESHRKELOREHSQKSSTKDVATVLKONISSKSTNNPK 773
DR	GO; GO:000233; F:peptidase activity; IEA.			Db 2058 PQVNHSQINESHRKELOREHSQKSSTKDVATVLKONISSKSTNNPK 2110
DR	GO; GO:0004289; F:subtilase activity; IEA.			RESULT 4
DR	GO; GO:000550; P:protein analysis and peptidolysis; IEA.			Q8DQP7 ID Q8DQP7 PRELIMINARY; PRT; 2144 AA.
DR	Interpro; IPR01899; Gram_pos_anchor.			PFam; PF00246; Gram_pos_anchor; 1.
DR	InterPro; IPR003137; PA.			PFam; PF00082; Peptidase_S8; 1.
DR	InterPro; IPR00209; Pept_S8_553.			PFam; PF05922; Subtilisin_N; 1.
DR	Interpro; IPR010259; Prot_inh_S8A.			
DR	Interpro; IPR01680; WD40.			
DR	Pfam; PF06280; DUF1034; 1.			
DR	Pfam; PF00246; Gram_pos_anchor; 1.			
DR	Pfam; PF00082; Peptidase_S8; 1.			
DR	PFam; PF05922; Subtilisin_N; 1.			



KW	Hypothetical protein.	Db	1933 KEELGKKNIS-----KDITNLLEITTTKKKNPISKSNN 1966
SQ	SEQUENCE	SEQUENCE	
Query Match	6.7%; Score 270.5; DB 2; Length 2649;	RESULT	6
Best Local Similarity	20.6%; Pred. No. 0.0052;	ID	081B84
Matches	206; Conservative 160; Mismatches 327; Indels 307; Gaps 49;	PRELIMINARY;	PRT; 1642 AA.
QY	1 KUGBIABSKFKQLGN---GKEGSLKKDTGTGVHHHOENESIKEKSFTIDRNISTIRD 56	AC	081B84;
Db	1044 KONNLIERNYRUDNHLHQVNRAKONNDISIINYNSKGSTGTVQNRSVTTEFIRS 1103	DT	01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
QY	57 RR-----NKLKLKLIKKEPREVDDFTSFTGKRMEEVDYKDDGNTIAYDD---- 102	DT	01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
Db	1104 FBDKIKLKLQNERLKKEKIEKDEKEVKVNDY-KKMEKIKKMQD--NLFBATDKHEK 1159	DE	Hypothetical protein MAL8P1.29.
QY	103 -GTDLEYETEKEDEIKSCKIVGLSPSDGHDFFLTGKISNSVSKNAKYVGN--NYKSTEIK 159	GN	Name=MAL8P1.29;
Db	1160 LACILEMNSKOMBLIKIN-----KDKDMLTIELESQICN-NYASNEHLANTNUDV 1211	OS	Plasmodium falciparum (Isolate 3D7).
QY	214 KETKSEPYVSSYGNVILGEGDLSKPKPDNL----- 245	OX	Bukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
Db	1272 KJLNMKGENINYEYERSIDKLKSIDKLUKEQONLKLTEQICDLKERNIMLEKAQLRNDBSS 1331	RN	[1]
QY	246 -----TKOESG-----KYS----- 256	RA	SEQUENCE FROM N.A.
Db	1332 NTTSITSDGTTINNEKIMKBEJEALYKDKCLKLKSNLEEKNKNTINTLTKTSEQS 1391	RL	Quail M., Barrell B., Harris D., Berriman M., Pain A., Hall N.,
QY	257 --SEKOYLKLONILRKGYALKVTYNGPKTDMLBECNGVSKEDIAKIQKANPRL 313	DR	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
Db	1392 ITFLANKKIKCLKENKNIQKNEYSI-----NDLKLNKIQD-EDITKENSIPSI- 1442	EMBL; AL844507; CADB1123 1;	.
QY	314 SERTTYADSRNVEDGRSTQSVMSALD--GENIR--YQVTFPKMDKG----EAD 362	KW	Hypothetical protein.
Db	1443 --TTFDLANNEKLEHNTWKQDTESTQIDDSDYDMARIKKYDISHKONSENCYDQIKTBID 1500	SEQUENCE	1642 AA; 194753 MW; 701F5D28B2BB8097 CRC64;
QY	410 NILSMOKNYPNPSKSNKTYVNPFPY-----LRGKTSKGG---FNWLVRNESVVD 457	Query Match	6.6%; Score 265.5; DB 2; Length 1642;
Db	1558 NNRKENITYNLLKENKVNPNVNFEDSRANLQNNESANGNESPNYEVNNNE--H 1614	Best Local Similarity	20.4%; Pred. No. 0.0048;
QY	1501 KONYKEKCTPTNKEQTDTNLNTLYSDVKKEMTENODDYN-----ILDDNKLINETNSMF 1557	Mismatches	313; Indels 297; Gaps 46;
Db	1416 IKSXKIVGVLSPSKDGHDFFLTGKISNSVSKNAKYVGNKSYBIAKMYKDFTKWTFLY 175	Matches	190; Conservative 133; Mismatches 313;
QY	176 ANINDIVGLAFAGDMFLFKVNDQKKAELKIRMPPEKKEKSEPVYSSYGNVILGEG 235	Db	104 ELNNRNRRNRRNRRNRRNNSNSFIMDEKEKUNTRNMLKHKRKYDPYNTS----DEK 158
Db	202 M-----DNKSSNNYNSGK----- 275	QY	61 DIKKLKRKFREVDDFTSETGKRM---EVDYKIDKGKNTIAYDDGTDLEYETEKLDE 115
QY	176 ANINDIVGLAFAGDMFLFKVNDQKKAELKIRMPPEKKEKSEPVYSSYGNVILGEG 235	Db	159 NKIR-----KKLNDIINKYKGMIIPSYEEENDKYSKNSGSFNFNISDD-----DE 201
Db	236 DKNN-----EMKEY----STDSDIDURVESISHSKE----NNNIDKLKCH 275	QY	116 IKSXKIVGVLSPSKDGHDFFLTGKISNSVSKNAKYVGNKSYBIAKMYKDFTKWTFLY 175
QY	236 DISKNDKDLTNTKNSGKYSIDSEKQQYLLKONILRKGYA----- 275	Db	202 M-----DNKSSNNYNSGK----- 275
Db	276 NLJRNKQSNLVNFDRK-----RKENIFELANYMKKGKENGKNTNINSYDIENTHDKDK 330	QY	176 ANINDIVGLAFAGDMFLFKVNDQKKAELKIRMPPEKKEKSEPVYSSYGNVILGEG 235
QY	276 -----LKVTYNGK-----TDMLEGN-----GVSKE-DIAKIQKANPRLS 314	Db	116 IKSXKIVGVLSPSKDGHDFFLTGKISNSVSKNAKYVGNKSYBIAKMYKDFTKWTFLY 175
Db	331 NEILNNLDINSYNEKBLCSYNNMKENDFNKIGISKOPNDVKSERKYNKYEKDHNLYNLS 390	QY	236 DISKNDKDLTNTKNSGKYSIDSEKQQYLLKONILRKGYA----- 275
QY	315 ETIYADSRNVEDGR-STQSVMSALDGFNTRYQVFFKNDK----- 357	Db	202 M-----DNKSSNNYNSGK----- 275
Db	391 DGKSKIRTNTVDDNSSSHNTSLNL---ENESNFMIDLEKDKRDKEKEVKEKLSLRRNS 448	QY	176 ANINDIVGLAFAGDMFLFKVNDQKKAELKIRMPPEKKEKSEPVYSSYGNVILGEG 235
QY	358 -----GEATKDO-GNLVTD-----SSKLVLFGKDDKETVGEDKFNVFAIKE 397	Db	236 DISKNDKDLTNTKNSGKYSIDSEKQQYLLKONILRKGYA----- 275
Db	449 ETYENEKEGETHKKOBELNKVLVLDMTKIDINEKKKKVLYKDYSE---NYERKISS 503	QY	176 ANINDIVGLAFAGDMFLFKVNDQKKAELKIRMPPEKKEKSEPVYSSYGNVILGEG 235
QY	542 KGNTSIEYADGKSVVFNINDK-----RNGFGEIQOHQIHYINGKEYTSFNDIKOIDK 595	Db	391 DGKSKIRTNTVDDNSSSHNTSLNL---ENESNFMIDLEKDKRDKEKEVKEKLSLRRNS 448
Db	1720 NHNSNIKSD-----NVNKVTECLNKINNEFANNESTLNNTETERNSTNDLNKIYE 1773	QY	358 -----GEATKDO-GNLVTD-----SSKLVLFGKDDKETVGEDKFNVFAIKE 397
QY	596 TNIKI---VVKOPARNNTVKEFTLN-RDTGESELKPHRTVUTIONGKEMSTTVE-EE 650	Db	449 ETYENEKEGETHKKOBELNKVLVLDMTKIDINEKKKKVLYKDYSE---NYERKISS 503
Db	1774 DNNIAYNNKIKIENY-NDODLKNYLNNSOKNNNNHEKND--NNTNEKKKKGEAWIDKN 1831	QY	449 ETYENEKEGETHKKOBELNKVLVLDMTKIDINEKKKKVLYKDYSE---NYERKISS 503
QY	651 DFLPVPYKGELEKGQFGWELSGFEGKDDGTVINUSKDFIKPWFVKRKEKEENKP 710	Db	398 DESMLFIDTPKVNLSKMKYFNSKSKVYVNPFPYFLRGKISKDFGPFWNLVRNESVVD 457
Db	1832 NEVLPYTK---IECVCVLSD-----EKEAGTKRKNK-----KRSCKRKNST 1872	QY	504 EN---IDIGPTNRSFLVGDGNNTKDMWHINENDR-----DNNDN----INN----D 544
QY	711 -----TPDVK-----KRDQFQVN-H-----SQLNESH 733	Db	458 NYLIGLHDIDTRDFTKLNKVLGMDWMGKMDYKANGFDFKVTDGNY-----YL 510
Db	1873 VNNMRNRNTYIVKRPSESIMKMKTEFANKKKKNPLHMKTNGLDSLVNDISKLVNKI 1932	QY	545 NTINNDNINNNVNVNK---ESNEDF-----YKANKUSKEKDNVONTNKKIKKS 596
QY	734 KEDQREEHQSQSDSTKQVTT--ATVLKNNSKSTNN 770	Db	511 QTCYSDJMAKAVGHVYFLYDNV-KEPVNIDPKGN---TSEYADGKSVVFNINDKNN 565
QY	566 GFDGEIQBOHTYINGKEYTSFNDIKOJIDKTLNKIVVKDFARNTTVREFLINKDQGEGV 625	Db	597 NSGNYNDAN-----LSQDISNBANLKVNTNSMDYTKYDDK-----GINTKN 642

Db	643	TP----KKSAYDKNKGLLNDTKLMIGHNN---QKENVIRND-EKVINGINNSPK	691	Db	1648	YEGSKNUISDNDVETNISDINTNKQDQ- EVESKRIFETNDNINGHISSDNNKINKMK	1705
Qy	626	EUKPHRVTYTO---NGKEMSTIVSBEEDTFPLPVYKGELKGYOFDGWEISGFEDKD-	680	Qy	265	KDNITLRKGYALKVUTYNGPKTD-MLEGNGVYSKDIAKQANPFLR----ALSETTI	318
Db	692	DJDNKRNIDCIOVNANNNNNNNNNTFIDAMEAISGNDEQYKNVQED-HVJMGVEENKN	750	Db	1706	QNNLINESQDKNIDVH--KLKLUKNERHTVSDESLIKREENGNTTRKNGSINNEKI	1763
Qy	681	-----AGVYINISKDTPI----KPVF--	697	Qy	319	YADSRWV-----EDGRSTQSVLMSALDGEMIRYQVFIFKN-----	355
Db	751	MOTNCSSNNKNYINNDRWNLILKEKFDMMMKPGVUDNITINKENELVLOQNEKKDIFNK	810	Db	1764	EERKENVNNDTETIGKREENTESDDIKIOSKTSNENINNKNTLYTDNYKDKSYNAQGTHG	1823
Qy	698	-----KKIEEKKEEENK--PTFDVSKKKDNPNQWH-----SQLNESH---RKEDLQRE	740	Qy	356	-----PKGEAID---KDGSLV-----	368
Db	811	SHGNEKEPLDNNKNVKNVILKQVKNDIHDYKVSNIGNEIDENRKKKENLNS	870	Db	1824	ENDETNTGNTISNDGDKNVIQDQVSKGENIQLQENKEDILPSVTINSLGDKVYKELNS	1883
Qy	741	EHSQSKSTDKO-VTATVLDKNNISKSSTNNP	772	Qy	369	-TDSSKVLVFLGKGDKEYTGDFKNV---EAIKEGSMFLDTK-----PVNLS-	412
Db	871	TNQEKGIGKNIKLNKTSYSKNNLSSYAEKPH	903	Db	1884	PEDIKMEVAHKNIQNTSEDELGTOQKDNERNKEDDKSPNGVEENHQENDKIGVUNLS	1943
OS		OS		Qy	413	MDKNVFNPSKSNKTYVRNPFLYRK- ISDKGGFWMLR--VNESVDNY-----L	460
OC		Bukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		Db	1944	MNNNESTIGNSDTI--NOHLINEGKNTIHKGVNNTSETNEMTNNGTQNTISNEQFEKNI	2000
RN		[1]		Qy	461	TYGLDHIDNTDFNKNVKNVKGDDMWGMKDYANGFPDKVTDMDGNYVYQTGVSIDLNAK	520
RP		SEQUENCE FROM N.A.		Db	2001	IRGD-DIKDRMENVKELEDEFGN-----NINKNDNAKILMELIKNQGTDSDAD	2052
RX		MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;		Qy	521	AVGVHYQFLYQVNUK-PEVNIDPKGNTSIEBADGKSVFVNINDKRNQFGBIBOHEYIYIN	579
RA		Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,		Db	2053	DISTNEDSKDQMDQIENKENBNTHKINNNVNE-----KOKI--SNDKEDNIVPPE-HKEDHIS	2105
RA		Carlton J.M., Pain A., Nelson K.E., Bowman S.G., Pauslins T.T., James K.,		Qy	580	GKEVTSFENDIQKIDKTLNKTIVKWDPPARTVTKERFLINKDTGEVSELKPHRVYTQNG	639
RA		Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Pertea M.S., Nene V., Shallom S.J., Suh B., Peterson J., Angulo S.,		Db	2106	DNKKKEFDNVNLEIPKGNH-----LDDKETITQVEEKSIGDKSMENNIVST--NDG	2157
RA		Peretea M., Allen J., Selengut J., Hafft D., Roos D.S., Ralph S.A.,		Qy	640	KEMSSTIVSERDFPLPVYKGELKGYOFDGWBISFEGKGDAGYVILSK-----	689
RA		McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,		Db	2158	KDIH--IQEIDI-----KEMIINNVNDKHSKONLHDEPNKVEEKKH	2204
RA		Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,		Qy	690	--DTFTKPVFKTKEKEENKPT---FDVSKKDNPOVNHSQNLNEHRKDQORE	741
RA		Fraser C.M., Barrell B.,		Db	2205	BIADHKDKKEKEIQDNDSKNEPSENINLVDNAQDDK-----NISKLTNDIHDQE	2256
RT		"Genome sequence of the human malaria parasite Plasmodium		Qy	742	HSQSKSTDKOVTATVLDKNNISKS	766
RT		falciparum";		Db	2257	KGTNNNDSSVVE-----HNVSDKT	2273
RL		Nature 419:498-511(2002).					
DR		DR EMBL; AE014822; AN37017.1; -.					
KW		Hypothetical protein.					
SQ		SEQUENCE 3504 AA; 408308 MW; B9454D48D55BB4FO CRC64;					
Query Match		6.6%; Score 264; DB 2; Length 3504;					
Best Local Similarity		21.2%; Pred: No. 0.014; Mismatches 196; Conservative 131; Gaps 346; Indels 252; Gaps 43;					
Matches							
Qy	11	KNLGNGKEGSJKKDITQEVHQQEN-BESTKEK----SSFTIDRNISTIRDFENKDL	62	RESULT 8			
Db	1432	KNSHENSEDNTIYTEEYKGKNSQPNENTDNTIPEKEGNGKNSQQNSHNTWKEKGENTNS	1491	ID Q7RG2	Q7RG2	PRELIMINARY;	PRP; 1850 AA.
Qy	63	KKLKIKKPREVDDFTBTKMEEYKYDKR--GNTIAYDDGDLVEYETKDEBK	118	ID Q7RG2	Q7RG2	PRELIMINARY;	PRP; 1850 AA.
Db	1492	QQ-----NSDHNITTEYEYKGKNSQENTDNTIMWEGYKGKNSKEKNSBEDIAS	1539	AC Q7RG2;	AC Q7RG2;		
Qy	119	KIYGV----LSPSKD-GHFELTIGKISNVSKVAKVYGG----NNYKSIERATAKDF--	165	DT 01-MAR-2004	DT 01-MAR-2004	(TREMBLrel. 26, Last sequence update)	
Db	1540	--YEIDKRNISHEENDQHDFYTESRINKEFHNDVNVISGDSNEKEGKSVNNSIEDIPD	1597	DT 01-MAR-2004	DT 01-MAR-2004	(TREMBLrel. 26, Last annotation update)	
Qy	166	-HSKMTFDLYANINDVGLAFAGDMRLFKNDOKKAKEKIRMPKEKETKE--YP	221	DE 01-MAR-2004	DE 01-MAR-2004	Hypothetical protein.	
Db	1598	GRGKNIQDDILLEEN-----GKSKFENIEBEDKISKTO-KSKISHEAAGHFTP	1647	GN Name-PfY4034;	GN Name-PfY4034;	Name-PfY4034;	
Qy	222	VVS-----SYGVNIELGEGDISKNPDNLTMESGKTY-----SDSEKOQYL	264	OS Plasmidium yoelii yoelii.	OS Plasmidium yoelii yoelii.	Plasmidium yoelii yoelii.	
		NCBI_TAXID=73239;		OX NCBITAXID=73239;	OX NCBITAXID=73239;		
		RN [1]		RN SEQUENCE FROM N.A.	RN SEQUENCE FROM N.A.		
		STRAIN=17XNL;		RC STRAIN=17XNL;	RC STRAIN=17XNL;		
		PubMed=12368805; DOI=10.1038/nature01099;		RC PubMed=12368805; DOI=10.1038/nature01099;	RC PubMed=12368805; DOI=10.1038/nature01099;		
		Carlton J.M., Angulo S.V., Suh B.B., Koo J.W., Pertea M.,		RA Carlton J.M., Angulo S.V., Suh B.B., Koo J.W., Pertea M.,	RA Carlton J.M., Angulo S.V., Suh B.B., Koo J.W., Pertea M.,		
		Silva J.C., Brnoiaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,		RA Silva J.C., Brnoiaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,	RA Silva J.C., Brnoiaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,		
		Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Biwell S.L.,		RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Biwell S.L.,	RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Biwell S.L.,		
		Shallom S.J., van Aken S.B., Riedmiller S.B., Feldblum T.V.,		RA Shallom S.J., van Aken S.B., Riedmiller S.B., Feldblum T.V.,	RA Shallom S.J., van Aken S.B., Riedmiller S.B., Feldblum T.V.,		
		Cho J.K., Quackenbush J., Seeger M., Shoabai A., Cummings L.M.,		RA Cho J.K., Quackenbush J., Seeger M., Shoabai A., Cummings L.M.,	RA Cho J.K., Quackenbush J., Seeger M., Shoabai A., Cummings L.M.,		
		Florens L., Yates R. III, Raine J.D., Sinden R.E., Harris M.A.,		RA Florens L., Yates R. III, Raine J.D., Sinden R.E., Harris M.A.,	RA Florens L., Yates R. III, Raine J.D., Sinden R.E., Harris M.A.,		
		Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B., Smith H.O., White O.R.,		RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B., Smith H.O., White O.R.,	RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B., Smith H.O., White O.R.,		
		van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,		RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,	RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,		
		Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,		RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,	RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,		





QY	730 ESHRKEDLOREHQSOKSDSTKDVTAT-VLDRNNNTSKSTNNP 771	QY	285 KUDMELGGGVVSE--DIAKTOCANPNRLRAMESETTYADSDNEDGRSTSVLMSALD-G 341	QY	1188 DSNDSSD-----SNKSIROSDQTQIKKKSSSSNKEKKP 1224	Db	765 KKLIEETDKSISBEEYKNTILKVKDVBYIKVHT----NELISFRNKQITLKDKLN 819
QTRMS4	PRELIMINARY;	PRT;	2661 AA.	QY	342 FNIIRY----QVFPEK----MNDGEAIDQGNYV----TDSSKLVL-----G 378	Db	820 INTVKENNSTDQYKTFNLTIDKKTTELDTKFTDVALNDEHSNNELMKYFNNLKLNG 879
ID	QTRMS4;			QY	379 RDKD---EVGDEDKNEAKEDGSMFLFDTPKPNLMSMOKYFNPFSKNCIVRPFY 434	Db	879 KKLIEETDKSISBEEYKNTILKVKDVBYIKVHT----NELISFRNKQITLKDKLN 819
AC	01-MAR-2004 (TREMBIrel. 26, Created)			QY	435 IRGKLSDKGGFENWELRVNESVVDNVLYGDIIADNTRDNFKLKNVKOGDIDMWGMKDYA 494	Db	880 RPKKKNLYQDGEKERAIDDKKKRADINENVKTEVAIYASINYNSDEMREIEKSIES 939
DT	01-MAR-2004 (TREMBIrel. 26, Last sequence update)			QY	940 INTQVIKE-----VKTNTNLNBIGKLUQYFHDGKGRENK----- 977	Db	940 INTQVIKE-----VKTNTNLNBIGKLUQYFHDGKGRENK----- 977
DE	Rhoptry Protein (Fragment).			QY	495 NGFPDKVTMDGNVYIQTGVSDLNAKAVGMYMOFLDVKREVEVNIDPKGMNISYEADGKS 554	Db	978 -YPDBINKRNEL-----KVVG----QQIDOHINKLEIKNNSGSXNDEIRGKT 1021
GN	Name=PY02104;			QY	555 WVF-WNDKRNGFDGEIQEHOIHYINGKEYTSFN DIKOIDKTNKIVKWDPARNTV 612	Db	1022 DKFENVYDKE-----TINYKDPKEFEKINI-VTKDEKENIYI 1059
OS	Plasmodium yoelii yoelii.			QY	613 KEFLINKDTGEVSELMKPHRT-----VTLONGKEMSTSIVSEEDFLPVVKGELEKGQ 666	Db	1060 K---INKLUDISEBEKNTSLKKWDINVSYQSGLSKGKLFEQID----- 1101
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			QY	667 FDGWEISGFCKKDGAYVIMSKDPTFIPV-FKKEEKGEEENKPTFD---VSKKDN 720	Db	1102 -----EKKGAGHMK-SMEYMDJDLNKKSQEENEMKINMDKMYINKMEA 1151
OX	NCBI_TaxID=73239;			QY	721 PQVNISQLNNSHRSKEDLOREHQSOKSDSTKDVTATWLKDN---ISSKSTNNP 772	Db	1152 IANISHDDDKQHH---NISKGEKEKISD-----IHKNSLKIQBPFSTSNIN 1194
PP	parasite Plasmodium yoelii yoelii;"			QY		Db	
RT	Genome sequence and comparative analysis of the model rodent malaria			QY		Db	
RL	Nature 419:512-519(2002);			QY		Db	
CC	- - - CAUTION: The sequence shown here is derived from an			QY		Db	
CC	preliminary data.			QY		Db	
CC	EMBL; AAC01000575; EAA21526_1; -.			QY		Db	
CC	InterPro; IPR011591; Pox_II_rel.			QY		Db	
CC	InterPro; IPR06499; ReticuloocyteBP.			QY		Db	
CC	Prodrom; PD04111; Pox_II_rel; 1.			QY		Db	
CC	TIGRFAMs; TIGR01612; 255KDa-fam; 1.			QY		Db	
CC	NON_TER 2661 2661			QY		Db	
CC	SQ00UNCE 2661 AA; 312269 MW; OBD02AD00712253B CRC64;			QY		Db	
Query Match	6.4%; Score 257.5; DB 2; Length 2661;			QY		Db	
Best Local Similarity	20.0%; Pred. No. 0.018;			QY		Db	
Matches	179; Conservative 151; Mismatches 294; Indels 271; Gaps 41;			QY		Db	
QY	22 KKDTGTGVEHHHQNEBESIK-EKSSEPTIDRNS-----TIRDENK-- 60			QY		Db	
Db	427 KKKELETTKLINKENESETVKELEIRRBLFKKYSEDEVAKKVVEELKKLKETIKOYINKE 486			QY		Db	
QY	61 -----DLKLKIKKFREVDFTETRGKNEBEYDQKYD-----KNNIADV 101			QY		Db	
Db	487 YIKQAIIDLKKRAIENNKKYIDELGKNTTPFOIEBYYKDKTYSTKSELSBIYKGNTIVELY 546			QY		Db	
QY	102 DGTDLBEYEBKLDKSK-----IYGUSPSKQGHFELGKSNVSKN---- 144			QY		Db	
Db	547 NEUSSVQNTTPIKIKKGKELETLSIDEVNPKQNMKBEVNLKQISETSEKNSLSN 606			QY		Db	
QY	145 ---AKVYVG---NNVKSIEKATDKDFHSKMTFIDLYANTINDIVGLAFAGDMRLFVKDN 198			QY		Db	
Db	607 ILETKVYIFGEIDN---DLNTKLGPKNKEQ-BLSNKIND-----YTKE 647			QY		Db	
QY	199 DQ-----KKAETKRMPEK-----KETKSEPYVSYGNVYIELGEGDSLK----- 239			QY		Db	
Db	648 DQLSWYQSKIEIRKHYNDODINVDTKGEAKQNDQSKYAKKIPIKEGETSKINEVT 707			QY		Db	
QY	240 -NKPDNLITNE-----SGKYSIDSERQVYKJDNIT---LRKGALKTYINPG 284			QY		Db	
Db	708 AMKKNELLRLKUKYIDFDGAYKEKAMSEHECTEANKKIRKIEBKSYLVNNYEKK---FNNs 764			QY		Db	
QY	- - - CAUTION: The sequence shown here is derived from an			QY		Db	
CC	preliminary data.			QY		Db	
CC	EMBL; AAC01000391; EAA20802_1; -.			QY		Db	
CC	InterPro; IPR011591; Botulinum; 1.			QY		Db	
CC	Prodrom; PDD001963; Botulinum; 1.			QY		Db	





RA Fraser C.M.; Barrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum";  
 RL Nature 419:498-511(2002).  
 DR EML: AE014841; AANZ5938.1; -.  
 DR HSSP; Q93IC2; INWT.  
 KW HYPOTHETICAL PROTEIN.  
 SEQUENCE 227 AA; 262841 MW; 6D5D4B8BF643339C CRC64;

Query Match 6.3%; Score 252.5; DB 2; Length 227;  
 Best Local Similarity 19.1%; Pred. No. 0.024;  
 Matches 190; Conservative 155; Mismatches 319; Indels 329; Gaps 49;  
 QY 10 FPKNLGNGKE-----GSSLKOPTGVH-----HHQENNESEIKEK 42  
 Db 690 YNNNNNNKETCTSINIKHSELKYPENKSHVNNSYMKNTNHLPHRNATTSNNRNEEYEK 749  
 QY 43 SSPTIDRNIST---IRDFFNK---DLKLII-----KKFREDDFT-----SE 79  
 Db 750 EK---DRNITNGNNNTIVEVNNSCIPPLKOMIPDGVRNSINKUNVNTQTSSVY 806  
 QY 80 TCKRMEEYDY-----KYDKKONTIADGGTDELYSTEKEUDIEKSIKYGVLSPSK 128  
 Db 807 TNKNDIDNSPPMPLINGIRESKVLSNNN----NGSIGFFSKLQNYHQSNNV----- 858  
 QY 129 DGHFETLGKISIVSKNAVKVYGNVYKSIIEIMAKYDPHSHKNTP-----DLYANIN----- 179  
 Db 859 -----NESYPLKNNMMKNN-----IEHNDKKNIFLVKNYEDTYSNIHNGI 900  
 QY 180 -----DIVDGLAFAGDMR-----LFVUDN-----DQKK----- 202  
 Db 901 HENSMILKVNLYNLLKACTPHGYSRHNOKMYTHEENTNINOKKSYHYHNGTTLKPLVNTNN 960  
 QY 203 -----AEIKIRMPPEKIKETKS---EVPVAVSGNVTELGEGLSKNPDMILKMEGSKI 253  
 Db 961 WVNFEADINLQAQKLUHSLSMGYEDKSMENYRNKLYNNNNNNNN-----NI 1013  
 QY 254 YSDSEKCOY--LLKONITLRKGYAL-----KVTTYNPGKTDMLEBNGVYSSKEDITAK--- 302  
 Db 1014 YDNNEYCQINNSYCFDHSIDLKQFPLNHQNSKULTHSNNNKNSFFNGINVESSHHLANPEI 1073  
 QY 303 -----IQKANP-----N 309  
 Db 1074 KTFAHNSYPILNQGLINCNPLOCIGDSNQRKHNVVYIKKNEYLNKNGSIINVLKREG 1133  
 QY 310 LRALS-----ETTYADSRNEDGRSTOSVLSMALDGNNIIRYQVFTRKND 356  
 Db 1134 LRKTSITHNGKESFSNMNDKVNVTMEGUNIQDNVANNNNKES-C-DN1KHMRLKSLFVSRE 1192  
 QY 357 K-GE--AIDKODNLVUTDSSKUFLFGK-DDKKEYTGDFKPNVAIKED-GSMMPID---TKP 408  
 Db 1193 SYGEHKSLDVOBCYVKNKL-I-NKVDKKT-E-DNNNNSYLNEEDDNASMQFYETNSNP 1248  
 QY 409 VNLSMDD-KYFNPSKSNKTVRNBEFYLRKGKISKGGEW-----LRNEST----- 455  
 Db 1249 YTDQENNMKVN---NVLNNNSNYYVDSKVNDSKAENKSDDILNNENIHTKD 1304  
 QY 456 -----VDPYLITYGDHLIDNTRDFNIKLNVIDGDDMW-GMKDYKANGFPDKV-TDMDG 506  
 Db 1305 QKKEKIQNNNTEFSEQADIENTRSQSEVYEEHEPU-WWVINASNEEKKSYELIVSDMSS 1363  
 QY 507 NYLQTCYSDIMAKAVCHYQFLYDNTPKPEVNIDPKNTSLEYADGSKSVVENDKRNG 566  
 Db 1364 NRVTNKYSDMNVNEVILNERDNLLTEKVKYOLEKENKMIDMVTVEENINTIKENTND 1423  
 QY 567 PFGCEIQRQHITYINGKEYTSFPIKQIDKTKNITKVKDFRANT-TVK--EFILANKDTG 622  
 Db 1424 INEEVR-----NEQKRESINHIND-----TMNINHITDEYNDTYPNIKOTECVHNENN 1472  
 QY 623 EVSELKEPHRVTTIONKEMMSSTIVSBRDFLPVYKGEBLEGKYQFGWESISGPEGKKDAG 682  
 Db 1473 MNNSIBQYTFVHDTRNNHLVK--NNQFPI-----FESEGLNLENFECK--V 1516

QY 683 YVINISKD-----TPIKPVFKCIEKKEBENKPF-----DYSKKKD 720  
 Db 1517 YIENNKKDDHGDSKTSNLTSRNICKSENNDHNEKENTIVVRKGEGKIKRKVSMKRN 1576  
 QY 721 PQVN-HSQLN-----ESRKEDLOREHSQSK 746  
 Db 1577 EKLNBENYINNITYDKMDHRQNDDITKRENDEEN 1609

Search completed: February 16, 2005, 19:19:52  
 Job time : 188 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: February 16, 2005, 19:11:30 ; Search time 44 Seconds

(without alignments)  
1690.354 Million cell updates/sec

Title: US-10-067-385-8  
Perfect score: 4026  
Sequence: 1 KLGEBIAESKFKNLGNKGEGS.....ATVLDKNNISSKSTNNPNK 773

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79;\*

1: pir1;\*  
2: pir2;\*  
3: pir3;\*  
4: pir4;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

## ALIGNMENTS

Result	No.	Score	Query	Match	Length	DB	ID	Description
1	4026	100.0	2140	2	F95074			serine proteinase, metalloproteinate, ORF MSV156 hypochthonioid prote
2	3709.5	92.1	2144	2	A797942			hypothetical prote
3	253.5	6.3	1127	2	T28317			hypothetical prote
4	250	6.2	1650	2	T8444			hypothetical prote
5	235.5	5.8	3724	2	T8427			hypothetical prote
6	231.5	5.8	2269	2	T8677			hypothetical prote
7	230	5.7	2401	2	T8676			hypothetical prote
8	228	5.7	4550	2	T8440			hypothetical prote
9	221.5	5.5	1711	2	T8429			hypothetical prote
10	219	5.4	1640	2	A74594			probable major surface protein
11	218	5.4	1631	1	SL2QK1			major merozoite surface protein
12	217.5	5.4	1639	2	S05603			major merozoite surface protein
13	217.5	5.4	1546	2	G96603			lipoprotein [m]ajor
14	215.5	5.4	1252	2	B42771			reticulocyte-binding protein
15	214	5.3	2339	1	A45597			DNA-directed RNA polymerase
16	212.5	5.3	1125	2	E90598			membrane nucleic acid
17	212.5	5.3	2166	2	G70163			hypothetical prote
18	212.5	5.3	2485	1	H71621			serine/threonine-protein kinase
19	212	5.3	1188	1	A71621			protein with 5'-3'
20	211.5	5.3	1104	1	A16866			microbial collagen
21	209	5.2	1558	2	B71603			RESA-H3 antigen
22	207.5	5.2	688	2	F82885			PP-hypothetical prote
23	207	5.1	1141	2	B88284			hypothetical prote
24	207	5.1	1191	2	B71622			chromosome segregase
25	207	5.1	1979	2	C71622			hypothetical prote
26	206.5	5.1	1622	2	A8717			probable cell surface protein
27	206	5.1	1526	2	A85605			mature-parasite-in
28	204.5	5.1	1516	2	B71619			RAD2 endonuclease
29	204.5		2500	2	G71609			hypothetical prote

GAF domain protein  
hypothetical prote  
fibrinogen-binding  
hypothetical prote  
superfamily I DNA  
repeat organellar  
hypothetical prote  
AMP-dept. acyl-coa  
hypothetical prote  
surface protein (L)  
hypothetical prote  
Ser/Thr protein kinase  
hypothetical prote  
hypothetical prote  
hypothetical prote  
surface-located membrane protein - Myc  
lmp1 protein - Myc

GAF domain protein  
hypothetical prote  
fibrinogen-binding  
hypothetical prote  
superfamily I DNA  
repeat organellar  
hypothetical prote  
AMP-dept. acyl-coa  
hypothetical prote  
surface protein (L)  
hypothetical prote  
Ser/Thr protein kinase  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
surface-located membrane protein - Myc  
lmp1 protein - Myc

Db	1694	IDKGNLVLTDSKLVLFGKDDKEYTSDKENVTEAKEDGSMFLIDTPKPVLNSMDKNYFNP	1753	Qy	301	AKIQKANPNAKALSETTIYADSNTEDGRSTOSVMSALQFNITRYQVETPKNDGEA	360
Qy	421	SKSNKIVRNPEFYLRGKISDKGGNWLRLNESVNDVLYIYGDLHIDNTRDPNKLNVK	480	Db	1638	AKIQKANPNLVLSSETIYADSNTEDGRSTOSVMSALQFNITRYQVETPKNDGEA	1697
Db	1754	SKSNKIVRNPEFYLRGKISDKGGNWLRLNESVNDVLYIYGDLHIDNTRDPNKLNVK	1813	Qy	361	IDKGNLVLTDSKLVLFGKDDKEYTSDKENVTEAKEDGSMFLIDTPKPVLNSMDKNYFNP	420
Qy	481	DGDIMWGMKDKYKANGFPDKTDMDGNYLQGTGYSIDLNAKAVGVAQVHQLFLDNVKPEVNID	540	Db	1698	IDKGNLVLTDSKLVLFGKDDKEYTSDKENVTEAKEDGSMFLIDTPKPVLNSMDKNYFNP	1757
Db	1814	DGDIMWGMKDKYKANGFPDKTDMDGNYLQGTGYSIDLNAKAVGVAQVHQLFLDNVKPEVNID	1873	Qy	421	SKSNKIVRNPEFYLRGKISDKGGNWLRLNESVNDVLYIYGDLHIDNTRDPNKLNVK	480
Qy	541	PKGNTSIEYAGKSVVFINDKRNNRFDGEQEHVYINGKEYTSENDIKOIDKTLNK	600	Db	1758	SKSNKIVRNPEFYLRGKISDKGGNWLRLNESVNDVLYIYGDLHIDNTRDPNKLNVK	1817
Db	1874	PKGNTSIEYADGKVVFINDKRNNRFDGEQEHVYINGKEYTSENDIKOIDKTLNK	1933	Qy	481	DGDIMWGMKDKYKANGFPDKTDMDGNYLQGTGYSIDLNAKAVGVAQVHQLFLDNVKPEVNID	540
Qy	601	IWKDFARNTVKEFLINKOTGESELKPHRTVTLONGKEMSTIVSEEDFLPVYKGE	660	Db	1818	DGDIMWGMKDKYKANGFPDKTDMDGNYLQGTGYSIDLNAKAVGVAQVHQLFLDNVKPEVNID	1877
Db	1934	IWKDFARNTVKEFLINKOTGESELKPHRTVTLONGKEMSTIVSEEDFLPVYKGE	1993	Qy	541	PKGNTSIEYADGKSVVFINDKRNNRFDGEQEHVYINGKEYTSENDIKOIDKTLNK	600
Qy	661	LEKGKQFDGWEISGFRRKQDGAVVINSKQDPIKPKLKEERKEBENKPFDFVKKDN	720	Db	1878	PKGNTSIEYADGKVVFINDKRNNRFDGEQEHVYINGKEYTSENDIKOIDKTLNK	1937
Db	1994	LEKGKQFDGWEISGFRRKQDGAVVINSKQDPIKPKLKEERKEBENKPFDFVKKDN	2053	Qy	601	IWKDFARNTVKEFLINKOTGESELKPHRTVTLONGKEMSTIVSEEDFLPVYKGE	660
Qy	721	PQVNHSQNLNEHRKEDLQREBHQSKSDSTKOVATVLDKNNISSKTTNNPK	773	Db	1938	IWKDFARNTVKEFLINKOTGESELKPHRTVTLONGKEMSTIVSEEDFLPVYKGE	1997
Db	2054	PQVNHSQNLNEHRKEDLQREBHQSKSDSTKOVATVLDKNNISSKTTNNPK	2106	Qy	661	LEKGKQFDGWEISGFRRKQDGAVVINSKQDPIKPKLKEERKEBENKPFDFVKKDN	720
<b>RESULT 2</b>							
A97942	metallopeptidase (EC 3.4.21.-) A [imported] - <i>Streptococcus pneumoniae</i> (strain R6)	C;Species: <i>Streptococcus pneumoniae</i>	C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004	Qy	721	PQVNHSQNLNEHRKEDLQREBHQSKSDSTKOVATVLDKNNISSKTTNNPK	773
C;Accession: A97942	R;Hoekings, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; Dehoff, B.S.; E;e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, B.J.; Lu, J.; Matsushima, P.; McAnren, S.; M;Y; P.; Sun, P.M.; Winkler, M.E.	J; Bacteriol. 183, 5709-5717, 2001	A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; A;Title: Genome of the Bacterium <i>Streptococcus pneumoniae</i> Strain R6.	Db	2058	PQVNHSQNLNEHRKEDLQREBHQSKSDSTKOVATVLDKNNISSKTTNNPK	2110
A;Reference number: A97872; MUID:21429245; PMID:1544234	A;Accession: A97942	A;Status: preliminary	A;Molecule type: DNA	Qy	721	PQVNHSQNLNEHRKEDLQREBHQSKSDSTKOVATVLDKNNISSKTTNNPK	773
A;Residue: 1-2144 <KUR>	A;Cross-references: UNIPROT:Q8DQP7; GB:AE007317; PIDN:AAK99365.1; PID:915458138; GSPDB:C;Genetics: C;Keywords: hydrolase; serine proteinase	Best Local Similarity 92.4%; Pred. No. 6.4e-151; Length 2144; Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;	Db	2058	PQVNHSQNLNEHRKEDLQREBHQSKSDSTKOVATVLDKNNISSKTTNNPK	2110	
<b>RESULT 3</b>							
Qy	1	KLGELTAESKFKNLNGKEGSKKDTGVEHHQHOBESTIKSSTIDRNISTRDPENK	60	Qy	7	ESFKKNLNGKEGSKKDTGVEHHQHOBESTIKSSTIDRNISTRDPENKDLK	63
Db	1339	KLGELTAESKFKNLNGKEGSKKDTGVEHHQHOBESTIKSSTIDRNISTRDPENK	1397	Db	196	BIEFRNDINQKETKINKKQDBLKLNKDLDESKKEPIKKQOEALKNTIDKKQOEALKLNKDEIN	255
Qy	61	DLKKLKKKPEREVDDTSETGKMERDYKDKGMIAYDGTOLBYEKKDKLISKI	120	Qy	64	KLIKKKPEREVDDTSETGKME-----EDYKDKGKNTIAYDGT-----DLEYETE	111
Db	1398	DLKKLKKKPEREVDDTSETGKMERDYKDKGMIAYDGTOLBYEKKDKLISKI	1457	Db	256	FNIDEKQKLQDQINSKINTLNENTIKGVNLNTETKNTKLNQNETLNKOSTIKSLDEQK	315
Qy	121	IYGLSLPSKDKGHFELIKLISNSKNAKVVYGSYKSLIEKATKVDHSKIMFLDLYANIND	180	Qy	112	KLDEIKSKIVGLSSKDKGHFELIKSNSKNAKVVYGSYKSLIEKATKVDHSKIMFLDLYANIND	161
Db	1458	IYGLSLPSKDKGHFELIKLISNSKNAKVVYGSYKSLIEKATKVDHSKIMFLDLYANIND	1517	Db	316	LIDELDKNTNNTISLYNSK---TKINTNQQLLESSLTDNNANI---NINELSKIK	367
Qy	181	IYDGLAFAGDKRLFVNDOKRAEKETKIRMPKIKETSEYFVYSSVGNVIEFLGDSLKN	240	Qy	162	KYDFHSKMTFDLYANINDVGLAFAGDKRLFV---DNUOKKA-EIKRMPKIKET-	216
Db	1518	IYDGLAFAGDKRLFVNDOKRAEKETKIRMPKIKETSEYFVYSSVGNVIEFLGDSLKN	1577	Db	368	LFDNDIQKLNDITEQNKITD--FFNNSTRIFKEKUDTEYKKDIDKNNLQKLEESYK	425
Qy	241	KPDNLTKMESCKYISSEKOQYLKOMILRKGYALKVTTNPQGKTDMLEGGNTYSKED	300	Qy	217	----KSEY---PYVSSVGNVIEFLGDSLKDLSKQSYKQVYKJLNNII	1637
Db	1578	KPDNLTKMESCKYISSEKOQYLKOMILRKGYALKVTTNPQGKTDMLEGGNTYSKED	1637	Qy	217	----KSEY---PYVSSVGNVIEFLGDSLKDLSKQSYKQVYKJLNNII	1637

Db 426 KIDEQTYKVKINKENKEINDIELKLNNLQKLEEEKKIDEQEYKKINKEY--NDIE 483  
 Qy 270 LRGKGALKVKTUNPG-----KTDMLLEGWVYSKEDIAKIQKAMPNLRASETTIYAD 321  
 Db 484 LKNNNLQKLEEEKKINKENDILKLNUDETSNTEFLNKLNISDFDKSREIAKLN--TEYEQ 541  
 Qy 322 SKN--VEDGRSTQSVL-----MSALDGFNIRYQVTFKMDKGKAIDKOGNLVTDSSK 373  
 Db 542 LRKDLENINKNKLNKLSNKLDSNKLSSB-----OLXSKNKL-BGIDKTNLSKRN- 592  
 Qy 374 LVLFGKODKEYTGBEDKVNVEATAKED--GSMFIDTKPVNLSMDKNYFN--PSKSNIY 427  
 Db 593 ---DKDEFYFSNEKDIIVVIENKIGNLSDIINKLNDQFKEVINSKIDSKNELS 648  
 Qy 428 VNRPEFY-----LRGKGSKGGFWELRVNFSWVNTLYQDHLHNTRD 472  
 Db 649 TMFDDIFNAKQIASITNNNIEN-SNKRD-----LNPFTISNEDSSKEL-LDBRK 698  
 Qy 473 FMIKLNKTDGMWDGMKYANGFPDKVTMDGNV-YLQTCYSDLNKA-----V 522  
 Db 699 YKQFD-KIKDAMTTEVKSF-E-NLQDIDSISKNMNLNTAYDINTKANDLDDKLNNY 756  
 Qy 523 GHVHQPLJDNVKE-----EVNIDPKGNTSIEFAD-GISWVFNINDKRNFGFBIGEQHI 576  
 Db 757 GBEPKNUYNNAASLDLTTIQKNDERVKOLNEYLERNQNSITNDIVN----- 805  
 Qy 577 YNGKEYTSFNDIKOIIDKLTNIKIVWDKFARNYTVEKFILMKDTGSEVSE--LKPHRVT 633  
 Db 806 FI--KELTKFNNIE-TNKSIA-ELLND--DINDKIFKLYKELNKISTNLKLYK- 855  
 Qy 634 VTIQNGKEMSTIVSEEDFILPVYGELEKGYQFDGWEISGEGKKDAGYVNLSLKF- 692  
 Db 856 NEIDNUNBKLSIVENLQFINSPLSTEFNQ-----SITH-----INFLNLTIA 900  
 Qy 693 -IKPVFKKIEKEKEBENKPTFDVSKKKDPQVNHQSOLNBSRKEDLORERHSQSKSTD 751  
 Db 901 GINDVLUKLNLKIMADTTRGDTNIR-----DETKNOISSENITSQKNEKNE--KD 950  
 Qy 752 VPAVV-----LKDNTNNSKSTTN 770  
 Db 951 LKKLISFNDKLNKCNISAGYIEYN 974

RESULT 4

T18444 hypothetical protein C0385c - malaria parasite (Plasmodium falciparum)  
 C;Species: Plasmodium falciparum  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C;Accession: T18444  
 R;Lawson, D.; Bowman, S.; Barrell, B.  
 A;Reference number: 218935  
 A;Accession: T18444  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-1650 <LAW>  
 A;Cross-references: UNIPROT:077328; EMBL:298547; NID:e1325376; PID:e1427940; PID:CAB111  
 C;Genetics:  
 A;Map position: 3  
 A;Intron: 1597/3; 1625/3  
 A;Note: C0385C

Query Match 6.2%; Score 250; DB 2; Length 1650;  
 Best Local Similarity 20.8%; Pred. No. 0\_0034;  
 Matches 205; Conservative 134; Mismatches 320; Indels 328; Gaps 48;

Qy 20 SLKKTGQVVERHHOBENESIKKEKSFIDRNSTIRDENPDKQKLLKRRKFERB--VTDFF 76  
 Db 394 SLMEVNEYERKIKYGRKBNKBYSNKBY-----HFKND-DSSIRKKNSSECQ 443  
 Qy 77 TSPGKR--MEEDYKVDDKGNTAY--DDGTDL--YETEKUDEJKSKIVGVLSSKD 129  
 Db 444 KKCKYKYLIEQKRYNFNDNNNAYKDDTHKKEKGYLYNMIVQSBYKYG--SNKTM 501

RESULT 5

T18427 hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)  
 C;Species: Plasmodium falciparum  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C;Accession: T18427  
 R;Lawson, D.; Bowman, S.; Barrell, B.  
 A;Reference number: 218935  
 A;Accession: T18427  
 A;Status: preliminary; translated from GB/EMBL/DDBJ

Query Match 6.2%; Score 250; DB 2; Length 1650;  
 Best Local Similarity 20.8%; Pred. No. 0\_0034;  
 Matches 205; Conservative 134; Mismatches 320; Indels 328; Gaps 48;

Qy 130 GHFETLQKIS---NWSKNAKVVYGVNYKI--EIKATKUDFHKMTFDLYANNDIV 182  
 Db 502 DEMEITWQHTMDFNINENLNKNIYF-DDEGYDPGEGKKKUDHLYTQKEYKNINDL 560  
 Qy 183 DGLAPAGDMRJFVKDQKKAIEKIRMPPEKIK-----ETKSEYPVVSYSGVNIELG-- 235  
 Db 561 K-----DHHLNDKETKEKONEIEEBEKKONIEEBEKKONIEEBEKKONIEEBEKK 614  
 Qy 236 ---DLSGKPKDLTKMSKESKQSYQVQLKDNII-----LR 271  
 Db 615 KKEIEBEKKKKI-EHEEKKKIDDEKKNTYANDKUTSHIDVNNCNTKIDALDHIEBEK 673  
 Qy 272 KQYALKVTTYNGKTD-----MLEANGVTSKEDIAK-----IQANPNL-- 310  
 Db 674 KTGKHEINLYKEIKNQYQMLNDENSIMLHEKQNTHQVNNLCOTDKMLOKENKLTN 733  
 Qy 311 RALESITTYAASNVRVEDGRSTSOSVLMALDGFTNITYQVFPTKMDGEAIDKGNLWTD 370  
 Db 734 NDKKKTPFLLSKRN-----TSVNLSSKPG-----TLSTKQUNATIKTKD--VTD 779  
 Qy 371 SSKLVLEG-----KDKB-----YTGDEK-NYEA-----IKEGSMJFIDTK 407  
 Db 780 NEKKCYVHDHRKDNTIRKRNKFINTYKGRNVAEVGESEVNKNKVNKGDNKVNKMVEN 839  
 Qy 408 P-----VNLSMDKNYF-----NFSKSNKIVYRPEF 433  
 Db 840 QCDNNNNVNEVKQGDENVNWKNIERSESSFMTHRKNTSTSPTDCAKNEQIKYVPHL 899  
 Qy 434 YLRGKLSDKGGFWELR--VNSBVVDNY-----LIGDHLIDN----- 469  
 Db 900 Y-SNKQEDDKKSIFLKMINENNIKKYDKKEKISTLEKKVFKWDVITNNDEKHSSK 958  
 Qy 470 -TRDENI-----KUN----- 478  
 Db 959 INDDFNNTIDKORGKNNPVLNRKKNETKILEKKSKIMKIONNLMKCCTNFPNNK 1018  
 Qy 479 -VKGDITMDGMKYANGFPDKVTMDGNVY----- 1019 GISSSTSSSSPKTFKQFCGITEKPKNLSNLYKTCIRNMVNDSIKLANDKADLYKDKKTS 1078  
 Qy 514 YSDLNARAVGVYQFLIDNVKEPVNIDPKGNTSIEADGKSVFVN-----INDKRNNGFDG 569  
 Db 1079 FNDINRANKGMNF-----KRDV-----PNKINMNDTKGR-VFNFVTLNNYRN--- 1124  
 Qy 570 EIQEQQHYYINGKEYTSFNDIKOIIDKLTNIKIVWDKFARNYTVEKFILKOTGEVSELKP 629  
 Db 1125 -----YIRSNK-----NNVK-----NGKIV-----GKKVWLKE--KONGSLHP 1156  
 Qy 630 HRTVTIQNGKEMSTIVSEEDFILPVYGELEKGYQFDGWEISGEGKKDAGYVNLSK 689  
 Db 1157 EGVTEADKFLNSNTYNDKLIBEKGKDIINERMEK-----YKNNK-MKYKL--K 1201  
 Qy 690 DTPIKPVFKCLEKKEBENKPTFDVSKKKDPQVNHQSOLNBSRKEDLORER--HSQKS 746  
 Db 1202 SNSIPPIKKERKSNDN-----DNKNDNINSNNINSNNNDKCLFLESKEKDVRVILKNN 1258  
 Qy 747 DSTKDVTAVTVDKNNNSKSTNNPK 773  
 Db 1259 NTIVNNNTMFRQNSNSDNNTISLKKR 1285



QY	567 EETTRIGINIESINTKVFKEK-----VKENVTNLANKIKEKLKHDFSDFGKEGNIKY 618
Db	481 -----DGIMDKMK-DYKANGPFD-----KVTDMDGDNV-----YLQGSPDLAN 520
QY	619 TDKIRKINDIMAVSQOIDOHINGLDDIQKSESYSEMKEQINKLEKVSNTESLNDE 678
Db	521 AVGHTQFLYDNVKFVN-----DPKGNTSIEYADGSKVVFVIN----- 560
QY	679 IDKIQQQTIVTKIDKKRNTEBENKLSEISKIEKONTSLKVKDNLSTCQNLNLFL 738
Db	561 --DKRNNGFGEIIOQHIVN-----SKEYTSNDIKOIKDTKLN-TKL----- 601
QY	739 QIDEKKKAENTIKSMAYIDLDNIKKKSQEIETEMDIKMDINKMDINKMELKISHDDKK 798
Db	602 -----VVKDPAINTTKEPLNKDGEVSELKPHRVYTQNGKE 641
QY	799 HDKSRHKENISDIYDKSSKTIQDPSRESPIND-INKLQKVNSESQNHSNDIN----- 851
Db	642 MSSTIVSEEDFILPVYQGELEGQYQFDGWETSGFEGKDDAGYVNIUSKDTPIKPVFKKE 701
QY	852 -----OCLENBVANIYNIKLKNIKKIKIDKV 877
Db	702 EKKBEENKPTFDVSKKKDNPOVNHSQLNESHRKEDQREHHSOKSDST--KDV----T 753
QY	878 BYTSBIEK----NICKNINDELBNNSBKVVKLEGDSLSKECRSKLNSTLDDKDIDECIKN 932
QY	754 ATVUDKONNISKSSTTN 770
Db	933 INVLKQNLNEETNTTN 949
RESULT 7	
QY	rhoptry protein - Plasmodium yoelii (fragment)
Db	T28676
C.Species:	Plasmodium yoelii
C.Acession:	T28676; A45521
C.Date:	15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
R.Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.	
Nol. Biotech. Parasitol. 76, 329-332, 1996	
A.Title: Comparison of two members of a multigene family coding for high-molecular mass	
A.Reference number: Z20507; PMID:97077455; PMID:8920022	
A.Accession: T28676	
A.Status: preliminary; translated from GB/EMBL/DBJ	
A.Molecule type: DNA	
A.Residues: 1-2401 <SIN>	
A.Cross-references: UNIPROT:O26216; EMBL:U36927; NID:91041784; PID:91041785; PIDN:AAH412	
A.Title: Identif. of a gene for a Plasmodium yoelii rhoptry protein. Multiple cd	
A.Reference number: A45521; PMID:9101660; PMID:2270106	
A.Status: preliminary	
A.Molecule type: DNA	
A.Residues: 2260-2401 <KEE>	
A.Cross-references: GB:N34281	
Query Match 5.7%; Score 230; DB 2; length 2401; Best Local Similarity 21.0%; Pred. No. 0.04; Gaps 46; Matches 187; Conservative 138; Mismatches 292; Indels 274; Gaps 46;	
QY	2 LGEIMESPKPKLGNGKEGSKKDTCTGVE-HEHQENBSIKEKSSFTIDRNISTIDPFENK 60
Db	589 IDKTYDKEPNILTDKTELETKFTGLSLNNHESNN--KELITYFYDLMKALGKONEN- 644
QY	61 DLKKLKKKREVDDSETSEIGRMERSDYKDDKGIA-----YDDGTDLIYET 110
Db	645 ---MLYKQFNEKE-----KADEVKKPDVKINVSNBITVSIYINNEDTENEI 693
QY	111 EK-----LDETKSKYGVLSPSKDHFBILGKLSNVSKNAVYGN 151
Db	694 GKSISLNLNTKLERKGAVNTNLNEKEKL-----KDYDPDFGK---EKNKCPDEN 742
RESULT 8	
QY	152 NYKS-IEIKATKDHSKMTMEDIYAN-----INDI---VGLGAFAGDMRLFLVKONDQKA 203
Db	743 KIKNDIDTLNOKDTSIETL-EIKKONSENHIDBIKGQDQLKKVPNKTW-NEDPKET 799
QY	204 ETKI-RMPBKIKETKSEPYVSSYQVNGECD--LSKNPNDLTMS-GKLY--- 254
Db	800 EKCLNTVEKIDCKNNTKEIDKLNBTSEKIEKDTSLEKLNQNLSTGCKSLGFLQOI 859
QY	255 -SDSEKOQYL-----DNI----- 275
Db	860 DEEKKAERTIKAMEAYIDLDNIKKKSQEIETEMDIKMDINKMDINKMELKISHDDKK 919
QY	276 LKTT--YNPQGKUDMLEBNGVVSKEDEIKAQKPNLRASETTYASRNVRNGRSTOS 333
Db	920 TYHTTSKNEHEKSDIRKNSLTIODFSRSBYINDIKBLEKNTVLESQRNTDINO--- 975
QY	334 VLMASLDG-FNITRYQVTFKMDKGAEIDLKGDNLVTDSSKLVLFGKDDKEYOB--- 387
Db	1011 KTKNAELNSNEKITOLKNNSSLKECOSK-IKSTDDNVSSECINN--ITNLKTYI-- 1063
QY	976 -YLSKLEMINYNL-----KLNKIKKID----- 1010
Db	388 DKFENF-----AIKEDGSMFLFIDTPKVNLSMDKQYFNPSKSNKTVRNPFPYLRK 438
QY	1064 VNKEKNINTYKNAEYBENYNVSINFNNTEMADTKSOYI-LNFK-----KONG- 1109
Db	498 PDKVTDMDGNVYLQTYGKDLNAKAVGVYQFLYDNVKPEVNIDPGKNTSIEYADGSKVVF 557
QY	1110 -----TNNDDYNIKELKEKK-----KSNVYKBAKGNTQCEKNEKK-- 1147
Db	558 NINDKRNNGFGEIODEIYINGKEYTSNDIKOIKDTKLN-KIVVKDFF--ARNTTVE 614
QY	1148 -----FEKYBQEVTLLN-KYAV--ELKNKFDFKTKNYSBQIKEKDIAHNTFTSQ 1195
Db	615 FLNKDGEVSEKPHRTITVQNGKENSSTIVSBEDRFLIPVYQGELEGQYQFDWEIS 674
QY	1196 --ADKSEKNEKNEKNEQRIEDEVAKKNSN-----KAILD----- 1229
Db	675 FEKGDAGYVNTISKDTPIKPVFKKE--KGEENKPTFDVSKKKDNFOV----- 723
QY	1230 -----IQSVEPP-KIKFKIKDQRTKSDDCKEKTJDITKISLISDQETKJ 1278
Db	724 -NHSQLNNSHQRKDLOREHHSQSDSTKDVTATWLDKNNISSKSTNNENK 773
QY	1279 ENKINLNTLEKLUESLKNQKNIEDQKEB----LDEVNSKKNIESNVQ 1324
Db	
RESULT 9	
QY	T18440 hypothetical protein C0425w - malaria parasite (Plasmodium falciparum)
Db	
C.Species:	Plasmodium falciparum
C.Date:	15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C.Accession:	T18440
R.Lawson, D.; Bowman, S.; Barrell, B.	
submitted to the EMBL Data Library, August 1997	
A.Status: preliminary; translated from GB/EMBL/DBJ	
A.Molecule type: DNA	
A.Residues: 1-440	
A.Cross-references: UNIPROT:O77336; EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CA	
A.Reference number: Z18935	
A.Status: preliminary; translated from GB/EMBL/DBJ	
A.Molecule type: DNA	
A.Residues: 1-445 <LAW>	
A.Cross-references: UNIPROT:O77336; EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CA	
C.Genetics:	
A.Map position: 3	
A.Note: C0425w	
Query Match 5.7%; Score 228; DB 2; length 4550; Best Local Similarity 21.6%; Pred. No. 0.11; Matches 193; Conservative 119; Mismatches 347; Indels 236; Gaps 43;	
QY	5 IAEISKPKLNGKE--GSLLKQTGV---EHHQNEESIKERKSSFTIDRNISTRD 56



A; Reference number: A24594; MUID:86014355; PMID:2995820

A; Accession: A24594

A; Molecule type: DNA

A; Residues: 1-1640 <HOL>

A; Cross-references: UNIPROT:P04933

C; Superfamily: major merozoite surface antigen

C; Keywords: surface antigen

Query Match 5.4%; Score 219; DB 2; Length 1640; Best Local Similarity 18.6%; Pred. No. 0.071; Mismatches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39; Matches 167;

Qy 23 KOPTG-VERHHQNEESIK-----EKSPTIDRNISIRDPBNKDKLICKFREVD 75

Db 232 KONGVKMEYDVKKKTENELIESBKKTIDKNNKATKEER-----KKLYAQOY 283

Qy 76 FTTSETGKRMEEVYKDDKGNNIAYDDGTDLSETEKUDRKSISGYVLSPSKDGFIL 135

Db 284 DLSIYNIKOLEE-----AHNLJSV-----LR---KRIDLK-----KNNENKEI 319

Qy 136 GKLNVSKNAKVYGGNNYKSTIEKATKTYDFH-----SKWTFDLYANINDIVGLAFA 188

Db 320 DKINEKNPPPANSNTNTLIDKNNKIEBHEKEIKEIAKTIKFNFNIDSFLTDPL----- 373

Qy 189 GDMRLFYVNDQKGAEKIRMPPEKIKERKSEIP-----YVSISGNV-----IELGE-GDISK 239

Db 374 ELEYYIREKNN-KNIDISAKVETKESTERPEYNGVTPYLSDINNALNBLNSFGDLI- 430

Qy 240 INKDNLTKMSKGKTSDEBKKQQYL-LKDNILL-----RGYALKVTTNPKGTDML 289

Db 431 -NPFDYTKEPSSKNIYDNEBKKKINEKEKIKIEKKKESDEKSYEYRSKLN-----DT 485

Qy 290 EG-----NGVYSKE-----DIAKIQKA-----NPNLRALESETTYADSRSRNVEDGRSTQSV 334

Db 486 KEYEKLLENEYDLSKFNNDLTLNFEKOMGKGRSYKVEKLHTPNTASYENSKHLKLT 545

Qy 335 LMSALDGE-----NII-----RYQVFTEKNDGKAEIDKGNNIYTDSKLUVLFKGDKREYT 385

Db 546 ALKYMEDSLRNIVKEVELKYYKLISKIEETLVE-----NIKKDEBOLF-----EKKIT 598

Qy 386 GEDKFVNVAIKEKGSMMLFDTKVNLMSMDKNNPKSNKSYKVNPERFYLRGKISDKGP 445

Db 599 KOENPKDPEKTLIEYSDIVKVQVQKVL-----MNKIDELKKTQOLIKNVE-----LGKINHVNSY 653

Qy 446 NWELRVNEMSVVDVNLITYGDLHONTDRNFIKL-----NVK----- 480

Db 654 KOENKQE-----PYYLIVLKKEIDKLKVFMKPVESLINEEKNKTOGOSDINSBPSTEGI 709

Qy 481 -----DSDIMDGWMKDYK----- 501

Db 710 TQGATTKPGQOQASALEGDSVQAQAOQAPPPVPUVPEAKAQVPTPPAPVNTKENV 769

Qy 502 TDMDGNYVLQTGQDSLNA-----KAVGHYQFLYDNVKPBNIDPKGNTSIEYADGKSVW 556

Db 770 SKLD-----YLEKLYQFLNTSYICHYIUVSHSTMNEKLUKOYKIKERBSKSCDPLI 826

Qy 557 FNIN-----DKRUNFGD-----EIQBQHIIYINGKEYTSFNDIKOIKDITNLKIVV 603

Db 827 FNIQONNIPWMSMEDSLANLSOLFMEYKEVNCLYKLKONDKIKULLEAKKVSTV 886

Qy 604 KDPARNTTKEFILANKDGTGEVSEIPLKPHPRVVTIONGKMSSTVSEBDPFLPVYKGILEK 663

Db 887 KTUSSSSMOPSLSTPDKPEVSANDDTSHSTNLANSLKLENLIS-----LGKNN 937

Qy 664 GKPQDFGWEISLGPQK-----KDAVGVIULSKDTPKPVFKKIEBKEEENKPTF-D 713

Db 938 IYQ-----BLIGQKSENPFYKILKSDTFYNESTPNFTVS KADDINSLANDESKRKLLRED 993

Qy 714 VSKEKGDNPWNHSOLNESHRKEDLQREEHSQKSDSTKOUTATLKDNNKISK-STINNP 771

Db 994 INKLKTIQLSFLDLYNKYKLKLERLFDKIKTVGKROMIKLILKEOLESKLNSLNP 1052

RESULT 11

SANQKL

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st

C;Species: Plasmodium falciparum

C;Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 09-Jun-2000

C;Accession: A25120

R; Mackay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.; B

EMBO J. 4, 3823-3829, 1985

A;Title: Polymorphism of the precursor for the major surface antigens of Plasmodium falci

A;Reference number: A91050; MUID:86136024; PMID:3004972

A;Molecule type: DNA

A;Residues: 1-1631 <MAC>

C;Comment: The merozoite stages of different strains have strain-specific surface antigens

C;C-Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The me

C;Superfamily: major merozoite surface antigen

C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane prote

F;1-19;Domain: signal sequence #status predicted <SIG>

F;20-1631;Product: major merozoite surface antigen #status predicted <MBN>

F;67-84;Region: 3-residue repeats (S-G-T-P)

F;1614-1631;Domain: membrane anchor #status predicted <MBN>

F;97,259,759,835,911,955,1049,1156,1165,1436,1563;Binding site: carbohydrate (Asn) (

Query Match 5.4%; Score 218; DB 1; Length 1631; Best Local Similarity 18.1%; Pred. No. 0.078; Mismatches 164; Conservative 162; Mismatches 339; Indels 240; Gaps 37; Matches 164;

Qy 23 KOPTG-VERHHQNEESIK-----EKSPTIDRNISIRDPBNKDKLICKFREVD 75

Db 232 KONGVKMEYDVKKKTENELIESBKKTIDKNNKATKEER-----KKLYAQOY 283

Qy 76 FTTSETGKRMEEVYKDDKGNNIAYDDGTDLSETEKUDRKSISGYVLSPSKD 135

Db 284 DLSIYNIKOLEE-----AHNLJSV-----LR---KRIDLK-----KNNENKEI 319

Qy 136 GKLNVSKNAKVYGGNNYKSTIEKATKTYDFH-----SKWTFDLYANINDIVGLAFA 188

Db 320 DKINEKNPPPANSNTNTLIDKNNKIEBHEKEIKEIAKTIKFNFNIDSFLTDPL----- 373

Qy 189 GDMRLFYVNDQKGAEKIRMPPEKIKERKSEIP-----YVSISGNV-----IELGE-GDISK 239

Db 374 ELEYYIREKNN-KNIDISAKVETKESTERPEYNGVTPYLSDINNALNBLNSFGDLI- 430

Qy 240 INKDNLTKMSKGKTSDEBKKQQYL-LKDNILL-----RGYALKVTTNPKGTDML 289

Db 431 -NPFDYTKEPSSKNIYDNEBKKKINEKEKIKIEKKKESDEKSYEYRSKLN-----DT 485

Qy 290 EG-----NGVYSKE-----DIAKIQKA-----NPNLRALESETTYADSRSRNVEDGRSTQSV 334

Db 486 KEYEKLLENEYDLSKFNNDLTLNFEKOMGKGRSYKVEKLHTPNTASYENSKHLKLT 545

Qy 335 LMSALDGE-----NII-----RYQVFTEKNDGKAEIDKGNNIYTDSKLUVLFKGDKREYT 385

Db 546 ALKYMEDSLRNIVKEVELKYYKLISKIEETLVE-----NIKKDEBOLF-----EKKIT 598

Qy 386 GEDKFVNVAIKEKGSMMLFDTKVNLMSMDKNNPKSNKSYKVNPERFYLRGKISDKGP 445

Db 599 KOENPKDPEKTLIEYSDIVKVQVQKVL-----MNKIDELKKTQOLIKNVE-----LGKINHVNSY 653

Qy 446 NWELRVNEMSVVDVNLITYGDLHONTDRNFIKL-----NVK----- 480

Db 654 KOENKQE-----PYYLIVLKKEIDKLKVFMKPVESLINEEKNKTOGOSDINSBPSTEGI 709

Qy 481 -----DSDIMDGWMKDYK----- 501

Db 710 TQGATTKPGQOQASALEGDSVQAQAOQAPPPVPUVPEAKAQVPTPPAPVNTKENV 769

Qy 502 TDMDGNYVLQTGQDSLNA-----KAVGHYQFLYDNVKPBNIDPKGNTSIEYADGKSVW 556

Db 770 SKLD-----YLEKLYQFLNTSYICHYIUVSHSTMNEKLUKOYKIKERBSKSCDPLI 826

Qy 557 FNIN-----DKRUNFGD-----EIQBQHIIYINGKEYTSFNDIKOIKDITNLKIVV 603

Db 827 FNIQONNIPWMSMEDSLANLSOLFMEYKEVNCLYKLKONDKIKULLEAKKVSTV 886

Qy 604 KDPARNTTKEFILANKDGTGEVSEIPLKPHPRVVTIONGKMSSTVSEBDPFLPVYKGILEK 663

Db 887 KTUSSSSMOPSLSTPDKPEVSANDDTSHSTNLANSLKLENLIS-----LGKNN 937

Qy 664 GKPQDFGWEISLGPQK-----KDAVGVIULSKDTPKPVFKKIEBKEEENKPTF-D 713

Db 938 IYQ-----BLIGQKSENPFYKILKSDTFYNESTPNFTVS KADDINSLANDESKRKLLRED 993

Qy 714 VSKEKGDNPWNHSOLNESHRKEDLQREEHSQKSDSTKOUTATLKDNNKISK-STINNP 771

Db 994 INKLKTIQLSFLDLYNKYKLKLERLFDKIKTVGKROMIKLILKEOLESKLNSLNP 1052

Qy 551 DGKSVVVIN-----DKRNGEDG-----BIOQHIIYINGKEYTSFNDIKOIKDIL 597

Db 812 DPLDLNLIQNNIPVMSMFSQLENLVYCEKVMVNLYKIKNDIPIKNULEAK 871  
 Qy 598 NIKIVVKDFARNVTKEFILNKDTGEVSELPKPHRTVTIQNGKEMSTIVSEEDFLPVY 657  
 Db 872 KVTSVKTLSSSMQPLSLTPDKPEVSANDTSHSNTLNNSLKFENILS----- 922  
 Qy 658 KGBLELKQYDFWEISPEGK-----KDAGYVNLSKDTPFKVFKKEKEENK 709  
 Db 923 LGRKNKIQY---ELIGOKSENFYEKILKOSDFTNEFSFTNFVKSADDINSLDESKR 978  
 Qy 710 PTP--DYSKKNPQVHNSQLNESHRKGEDLQREHQSOKSDTKDVTAVLUDKNNSSK-S 766  
 Db 979 KKLBEDINKLKLQLSFDLNNYKYLKLERLFDKKKTIVGKVQWQIKKLTKEQLESKLN 1038  
 Qy 767 TTNNP 771  
 Db 1039 SLNNP 1043

RESULT 12

S05603 major merozoite surface antigen precursor - malaria parasite (*Plasmodium falciparum*) (strain G3) #sequence\_revision 12-Feb-1993 #text\_change 09-Jul-2004  
 C;Species: Plasmodium falciparum  
 C;Accession: S05603; S04850  
 R;Myler, P.J.  
 submitted to the EMBL Data Library, April 1989  
 A;Reference number: S05603  
 A;Accession: S05603  
 A;Molecule type: mRNA  
 A;Residues: 1-639 <MYL>  
 A;Cross-references: UNIPROT:P04933; EMBL:X15063; NID:99896; PIDN:CAA33163.1; PID:99897  
 R;Myler, P.J.  
 Nucleic Acids Res. 17, 5401, 1989  
 A;Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plasmodium falciparum  
 A;Reference number: S04850; MUID:89345116; PMID:266888/  
 A;Accession: S04850  
 A;Molecule type: mRNA  
 A;Residues: 1504-1639 <MYL>  
 C;Cross-references: EMBL:X15063  
 C;Superfamily: major merozoite surface antigen  
 C;Keywords: glycoprotein; merozoite; surface antigen  
 C;Signal sequence: signal sequence #status predicted <SIG>  
 F;20-1639/Domain: signal sequence #status predicted <SIG>  
 Query Match 5.4%; Score 218; DB 2; Length 1639;  
 Best Local Similarity 18.6%; Pred. No. 0.079;  
 Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;  
 Qy 23 KOTTG-VERHHQHENESIK-----EKSPTIDRNISTIRPFENKLUKKKKFREDD 75  
 Db 232 KONVKGMDYKCNKPTENTNELBESKKTIDKKNKATKRE-----LGKLYQAQY 283  
 Qy 76 FISETGKMEEDYKIDKGKNAIYDGTDIYEETKJLDESKSKTIVGLSPSKDGFIL 135  
 Db 284 DLISIYNQKOLE-----AHNLISV---LG---KRIDTUK-----KNEKELL 319  
 Qy 136 GKITNSVSKNAVYGNVYKSTRIKATKDFH-----SKTMFEDYANINDIVGALFA 188  
 Db 320 DKINEIKOPPNSGNTPNLIDRKKIEERKEIARIKIKNFDLSFLDFPL----- 373  
 Qy 189 GDMRLFKDNQDKKAETKIRHPEKIKETKSEPV---YVSXGNV---IEGE-GDLSK 239  
 Db 374 ELEYVYLRKKN-KNIDISAKVETKESTEPEVNGVYPLSYNDINNALNHLNSFGDLI- 430  
 Qy 240 WPDNLTKMEESSKISYESKQYL-LKDNIL-----RGIAALKVTTNPCKDML 289  
 Db 431 -NPFDYTKEPKSNITDNERKPKFNEIKEKIKIEKIESKKSVDKSLN---DIT 485  
 Qy 290 EG----NGVVSKE----DTAKIOKA----NPNLRALSHPTIVADSRAVNTEDGRSTOSV 334  
 Db 486 KEYEKULNEIDISKPNNDILINFEROMGKRYKSYKVERKLHTFASYENSKHNLK 545

RESULT 13

G90603 Lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)  
 C;Species: Mycoplasma pulmonis  
 C;Accession: G90603  
 R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.; Nucleic Acids Res. 29, 2145-2153, 2001  
 A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis  
 A;Reference number: A99512; MUID:21267165; PMID:11353084  
 A;Accession: G90603  
 A;Status: Preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-1546 <CTIP>  
 A;Cross-references: UNIPROT:Q98P17; GB:ALA445566; PID:914090150; PIDN:CAIC13908.1; GSPDB:G  
 A;Experimental source: strain UAB CTIP  
 C;Genetic code: SGC3  
 A;Gene: MTRU\_7350  
 Query Match 5.4%; Score 217.5; DB 2; Length 1546;  
 Best Local Similarity 20.2%; Pred. No. 0.077; Mismatches 336; Indels 249; Gaps 40;  
 Matches 180; Conservative 125; Mismatches 336; Indels 249; Gaps 40;  
 Qy 1 KLGELAEASKFKLNGKNGEGSLK-----KDYGVEHHHQENRESIKEKSFRIDRNISTIRD 56  
 Db 103 KSKELAHSQCVTSVNEBNKFRLAKNTKNTKOPSKAONLDTISQDHLSKNSKNSN 162  
 Qy 57 FENKOLKLIKRFKREDDFTSETGKMEEDYKID-KDNILAIYDGTDIYEETKJLDE 115  
 Db 163 IKNEQSKKQLTANNKEEHSANLQTONILESENTQODLKEN-----NSNKEKELRN 214  
 Qy 116 IKSKLYGVLSPSKDKHFFETLGK-TSNVSKNAVYGNVYKSTRIKATKDFHSKMTFDL 174  
 Db 215 VLS-FNKQEAVKGSPEFKFKEVSKLESKINIDKVDTEITDKFADP----SFEQ 267

QY 175 YANINDIVDGLAFAGDMRLFVTKNDNQKKAIEKIRKMRPEKETKSYPWVSYGNVIELGE 234  
 Db :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 268 Y-IEKIKDOLFRMAST---IKDTYO-----TNKIFLINDSE----- 298  
 C;Species: Plasmodium vivax  
 C;Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 09-Sep-1997  
 C;Accession: B42771  
 R;Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.  
 Cell 69, 1213-1226, 1992.  
 A;Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.  
 A;Reference number: A42771; PMID:92315338; PMID:1617731  
 A;Accession: B42771  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-125; <GAL>  
 A;Cross-references: GB:MB8098; NID:9160627; PID:9160628  
 A;Experimental source: strain Belem, merozoites  
 C;Genetics:  
 A;Gene: RBP2

RESULT 14  
 B42771  
 reticulocyte-binding protein 2 - Plasmodium vivax (fragment)  
 C;Species: Plasmodium vivax  
 C;Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 09-Sep-1997  
 C;Accession: B42771  
 R;Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.  
 Cell 69, 1213-1226, 1992.  
 A;Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.  
 A;Reference number: A42771; PMID:92315338; PMID:1617731  
 A;Accession: B42771  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-125; <GAL>  
 A;Cross-references: GB:MB8098; NID:9160627; PID:9160628  
 A;Experimental source: strain Belem, merozoites  
 C;Genetics:  
 A;Gene: RBP2

RESULT 1

QY 737 LQREET-----HSQKSSTDKVATVLDKNNSKSTNN 770  
Db 2020 LEKEENIIVSEKDSVNOLKSEKKDIN---DDNNNNDDNNNN 2060  
Search completed: February 16, 2005, 19:21:34  
Job time : 51 secs